

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame plus p2n model

Run on: December 23, 2005, 18:47:32 ; Search time 4894 Seconds  
(without alignments)  
4169.759 Million cell updates/sec

Title: US-10-019-931-3

Perfect score: 1898  
Sequence: 1 MATSSMKSTIMAFPSMCH.....RLSALVDDKLGSGGLRRLS 359

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xip  
-Q=/cgm2.1/USPTO.spool.p/US10019931/runat.16122005.165218.18438/app.query.fasta\_1.519  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LIST=45  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pcc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10019931 @CGN 1 1 7415 @runat.16122005.165218.18438 -NCPU=6 -ICPU=3  
-NO MAP -LARGESUBSTRY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:.\*  
1: gb Da:.\*  
2: gb In:.\*  
3: gb Env:.\*  
4: gb Om:.\*  
5: gb Ov:.\*  
6: gb Pat:.\*  
7: gb Ph:.\*  
8: gb Pr:.\*  
9: gb Ro:.\*  
10: gb Sts:.\*  
11: gb Sy:.\*  
12: gb Un:.\*  
13: gb Vi:.\*  
14: gb Htg:.\*  
15: gb Pl:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	1898	100.0	1077	6	AX080458 Sequence
2	1898	100.0	1080	15	BT008847 Arabidops
3	1898	100.0	1273	6	CQ832234 Sequence

4	1898	100.0	1273	6	CQ881446	CQ881446 Sequence
5	1898	100.0	1273	15	AY084939	AY084939 Arabidops
6	1898	100.0	1347	15	AY099809	AY099809 Arabidops
7	1898	100.0	86380	15	AB010697	AB010697 Arabidops
8	1892	99.7	1270	15	AY054219	AY054219 Arabidops
9	1582	83.4	1236	15	AK117202	AK117202 Arabidops
10	1553	81.8	1041	6	AX080460	AX080460 Sequence
11	1553	81.8	1044	15	BT006241	BT006241 Arabidops
12	1519	80.0	1023	6	CQ832235	CQ832235 Sequence
13	1519	80.0	1023	6	CQ881447	CQ881447 Sequence
14	1076.5	56.7	115071	14	AC135566	AC135566 Medicago
15	1024	54.0	100108	15	AP008111	AP008111 Lotus cor
16	937	49.4	124300	15	AP004534	AP004534 Lotus cor
17	905.5	47.7	115362	15	AC149197	AC149197 Medicago
18	905.5	47.7	116174	15	AC148339	AC148339 Medicago
19	899.5	47.4	112154	15	AC148395	AC148395 Medicago
20	899.5	47.4	112154	15	AC146342	AC146342 Medicago
21	876	46.2	108265	14	AC119410	AC119410 Medicago
22	861.5	45.4	119355	15	AC144476	AC144476 Medicago
23	861.5	45.4	145905	15	AC147431	AC147431 Medicago
24	861	45.4	111277	15	AC150505	AC150505 Medicago
25	861	45.4	138586	15	AC144726	AC144726 Medicago
26	858.5	45.2	111277	15	AC150505	AC150505 Medicago
27	858.5	45.2	138586	15	AC144726	AC144726 Medicago
28	858	45.2	93941	14	AP007718	AP007718 Lotus cor
29	850.5	44.8	136553	15	AC149307	AC149307 Medicago
30	848.5	44.7	112917	15	AC123976	AC123976 Medicago
31	783.5	41.3	180062	14	AC151963	AC151963 Glycine m
32	776	40.9	10950	15	BT002108	BT002108 Arabidops
33	776	40.9	1154	15	AY087921	AY087921 Arabidops
34	776	40.9	1216	15	AY092961	AY092961 Arabidops
35	776	40.9	87543	15	AC006836	AC006836 Arabidops
36	770.5	40.6	996	6	AX506332	AX506332 Sequence
37	770.5	40.6	996	15	AF325065	AF325065 Arabidops
38	757	39.9	980	6	AX651508	AX651508 Sequence
39	757	39.9	981	6	AX412507	AX412507 Sequence
40	757	39.9	981	6	AX412570	AX412570 Sequence
41	757	39.9	981	6	AX507354	AX507354 Sequence
42	757	39.9	981	6	AX598857	AX598857 Sequence
43	757	39.9	981	15	AY113050	AY113050 Arabidops
44	757	39.9	1179	15	AF375458	AF375458 Arabidops
45	742	39.1	996	15	BT015380	BT015380 Arabidops

#### ALIGNMENTS

RESULT 1  
AX080458 1077 bp DNA linear PAT 22-FEB-2001  
LOCUS  
DEFINITION Sequence 1 from Parent W00102589.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Varin, L. and Gidda, S.  
Methods, compositions and genetic sequences for modulating  
flowering in plants, and plants genetically modified to flower  
early and tardily  
Patent: WO 0102589-A 1 11-JAN-2001,  
Varin, Luc (CA) ; Gidda, Satinder (CA)  
FEATURES  
source  
1. 1077  
/organism="Arabidopsis thaliana"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:3702"

ORIGIN  
Alignment Scores:

Pred. No.: 1.42e-156 Length: 1077  
 Score: 1898.00 Matches: 359  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-019-931-3 (1-359) x AX080458 (1-1077)

```

Qy 1 MetAlaThrSerSerMetLysSerLleProMetAlaIleProSerPheSerMetCysHis 20
Db 1 ATGGCTACCTCAAGCATCAAGAGCATTTCCAAATGGCGATCCCAAGTTTCTCCATGTGTGCAC 60
Qy 21 LysLeuGluLeuLeuLysGluGlyLysThrPheAspValProLysAlaGluGluAspGlu 40
Db 61 AAGCTCGAGCTCTTAAAGAAAGCAAACTCGCGAGTCGCCAAAGCCGAAAGAGATGAA 120
Qy 41 GlyLeuSerCysGluPheGlnGluMetLeuAspSerLeuProLysGluArgGlyTTPArg 60
Db 121 GGGCTAAGCTCGAGATTCGAAGAGATGTGGATTTCTTCTCCAAAGAGAGAGATGGAGA 180
Qy 61 ThrArgTyrLeuTyrLeuPheGlnGlyPheTTPCysGlnAlaLysGluIleGlnAlaIle 80
Db 181 ACTCGTTACCTTTCCTTACCTTCCAAAGGGTTTGGTCCCAAGCCAAAGAGATTCAAGCCATC 240
Qy 81 MetSerPheGlnLysPheGlnSerLeuGluAsnAspValValLeuAlaThrIlePro 100
Db 241 ATGTCTTTCCAAAAACATTTCCAAATCCCTCCAAACGAGTCGTTCTCGCCACCATACCT 300
Qy 101 LysSerGlyThrThrTTPLeuLysAlaLeuThrPheThrIleLeuAsnArgHisArgPhe 120
Db 301 AAATCCGGTACCAACTGGCTAAAGCTTTAACTTTACCATCTTAAACCGTACACGGGTTT 360
Qy 121 AspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHisAspLeuVal 140
Db 361 GATCCGGTGGCTCGAGTACCAACCAACCTCTTTCACATTCCAACCTCATGACCTTGTGA 420
Qy 141 ProPhePheGluTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAla 160
Db 421 CCTTTCTTCGAGTACACAGCTTTACGCCCAACGAGATGTTCGCCATCTTCGGGTCTAGCC 480
Qy 161 SerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGluThrIleGluLys 180
Db 481 AGTCCAAAGAACGTTTCGCAACCACTTACCGTTCCGTTCCCTAAAGAAACGATCGAGAAA 540
Qy 181 ProGlyValLysValValTyrLeuLysArgAsnProPheAspThrPheIleSerSerTTP 200
Db 541 CCCGGTGGAGGTCGTGTACTGTGCGCGAACCCTTTGACACATTCATCTCTTCGTGG 600
Qy 201 HisTyrThrAsnAsnIleLysSerGlySerValSerProValIleLeuAspGlnAlaPhe 220
Db 601 CATTCACCAACCAACATCAATTCGAGTCACTGAGCCAGCTTCTGTCAGACCAAGCTTTT 660
Qy 221 AspLeuTyrCysArgGlyValIleGlyPheGlyProPheTTPGluHisMetLeuGlyTyr 240
Db 661 GATCTGTATTGCGCGGAGGTGATCGGGGTTTGCGCCGTTTGGGAAACATGATGGGATAC 720
Qy 241 TrpArgGluSerLeuLysArgProGluLysValPhePheLeuArgTyrGluAspLeuLys 260
Db 721 TGGAGAGAGACTTGAAGAGACCAAGAAAGCTTCTTTTAAAGGTACGAGAGATCTCAAA 780
Qy 261 AspAspIleGluThrAsnLysArgLeuAlaThrPheLeuGluLeuProPheThrGlu 280
Db 781 GACGACATCGAGACCAACCTTGAAGAGCTTTCGACATTTCTTAGAGCTTCTTTCACCGAA 840
Qy 281 GluGluGluArgLysGlyValValLysAlaIleAlaGluLeuSerPheGluAsnLeu 300
Db 841 GAAAGAGAAACGAAAGGAGGTGTGAAGGCTATCGCCGAGCTGTGTAGCTTGAAGAACTTG 900
Qy 301 LysLysLeuGluValAlaLysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeu 320
Db 901 AAGAAGTTGGAGGTGAACAGTCAAAAGTCGATCAAGAACTTTGAGAAATCGATTCTTG 960

```

```

Qy 321 PheArgLysGlyGluValAspAspTTPValAsnTyrLeuSerProSerGlnValGluArg 340
Db 961 TTTCGAAAGAGCAAGAGTGAAGTGGAGTTTGAGTTTACTATTGTTCACCTTCAAGTGGAAAGA 1020
Qy 341 LeuSerAlaLeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArgLeuSer 359
Db 1021 TTGTCAAGCTTAGTGATGACCAAGTTAGTGGTCAATCTGTCTCACTTCAGGTTAGC 1077

```

# RESULT 2

```

LOCUS BT008847 1080 bp mRNA linear PLN 13-JUN-2003
DEFINITION Arabidopsis thaliana At5g07010 mRNA, complete cds.
ACCESSION BT008847
VERSION BT008847.1 GI:31711859
KEYWORDS FL1 CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

```

## REFERENCE

```

AUTHORS Kim,C.J., Chen,H., Cheuk,R., Shinn,P., Bowser,L., Carninci,P.,
Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
Wu,H.C., Yamada,K., Yu,G., Shinzaki,K., Davis,R.W., Theologis,A.
and Becker,J.R.
TITLE Arabidopsis ORF clones
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1080)
AUTHORS Kim,C.J., Chen,H., Cheuk,R., Shinn,P., Bowser,L., Carninci,P.,
Dale,J.M., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
Wu,H.C., Yamada,K., Yu,G., Shinzaki,K., Davis,R.W., Theologis,A.
and Becker,J.R.

```

## TITLE

JOURNAL

## COMMENT

Submitted (13-JUN-2003) Salk Institute Genomic Analysis Laboratory (SIGNAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA: 'RIKEN Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinzaki,K.

The Salk, Stanford, PGSC (SSP) Consortium members constructed and sequenced the pUNI (ORF) clones using the RAPL cDNAs: Kim,C.J., Chen,H., Cheuk,R., Shinn,P., Bowser,L., Chan,M.M., Chang,C.M., Dale,J.M., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K., Yu,G., Davis,R.W., Theologis,A., and Becker,J.R.

Kim,C.J. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinzaki,K. (RIKEN GSC) and Becker,J.R. (SSP/Salk) contributed equally to this work as PIs.

## FEATURES

source

```

1..1080
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="U25564"
/ecotype="Columbia"
/note="This clone is in pUNI 51"
1..1080
/note="steroid sulfotransferase-like protein"
/codon_start=1
/product="At5g07010"

```

CDS

/protein\_id="AAP68286.1"  
 /db\_xref="GI:31711860"  
 /translation="MATSSMKSIIPNAIPESMCKELKEKGTDPVPKAEDEGLSC  
 EFQEMLDLSPEKGRWRTLYLFOGFWCAGEKIQALMSQKHEQSLQENLVATIKS  
 GTTWLALPTFTILNRHFDVASTNPLFTSNPHDLVVFEEKLYANGVDPDLSGLA  
 SPRTFAHLPFGSLKETIEKPGVVYLCNPDFTISGHWYNNLSKSSVSVLDDQ  
 APFLYGRGYTGFPGFWEHMGVRESLKKPEKVFLERYELKDDITNLKRLATFEL  
 PPTREERKGVYKAIALCLSFENLKKLELVNKSISIKNPNRLPKRGVSDMVNVLIS  
 PSQVERLSALVDKLGSGSLTFLRS"

## ORIGIN

## Alignment Scores:

Pred. No.:	1.43e-156	Length:	1080
Score:	1898.00	Matches:	359
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	15	Gaps:	0

US-10-019-931-3 (1-359) x BT008847 (1-1080)

QY 1 MetAlaThSerSerMetLySerIleProMetAlaIleProSerPheSerMetCyHis 20  
 DB 1 ATGGTACCTCAAGCATGAAGACATTCATGCGATCCCAAGTTCTCATGTGTAC 60  
 QY 21 LysLeuGluLeuLeuLysGluGlyThrArgAspValProLysAlaGluGluAspGlu 40  
 DB 61 AAGCTGAGCTCTTAAAGAGCAAAACTCGGACGCTCCGAAAGCCGAAAGAGTGA 120  
 QY 41 GlyLeuSerCySgluPheGluGluMetLeuAspSerLeuProLysGluArgGlyTTPArg 60  
 DB 121 GGGCTAAGCTGCAGATTCCAAGAGATGTTGATTTCTTCTTAAGAGAGATGAGA 180  
 QY 61 ThrArgTyrLeuTyrLeuPheGluGlyPheTyrCySgluAlaLysGluIleGlnAlaIle 80  
 DB 181 ACTCGTACCTTACCTAATCCAAAGGTTTGGTGCCAAAGCAAGAGATTCAAGCATC 240  
 QY 81 MetSerPheGluLysHisPheGluInsLeuGluAsnAspValValLeuAlaThrIlePro 100  
 DB 241 ATGCTTTCCAAACCAATTCCTCAATCCCTGAAACGACGTCCTCGCACCATACCT 300  
 QY 101 LysSerGlyThrThrTyrLeuLysAlaLeuThrPheThrIleLeuAsnArgHisArgPhe 120  
 DB 301 AAATCCGATCAACCTGCTTAAAGCTTAACTTCACTTCACTTAACTTCACTTCA 360  
 QY 121 AspProValAlaSerSerThrAsnHisPProLeuPheThrSerAsnProHisAspLeuVal 140  
 DB 361 GATCCGGTTCCTCGAGTACCAACCAACCTCTTTCACTTCAACCTCATGACCTGTA 420  
 QY 141 ProPhePheGluTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAla 160  
 DB 421 CCTTTCTTCAGATCAAGCTTTACGCGCAACGAGATGTTCCCAATCTCTCGGCTAGCC 480  
 QY 161 SerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGluThrIleGluLys 180  
 DB 481 AGTCCAAAGACGTTCCGACCACTTACCTTCCCTTAAAGAAAGATCGAGAAA 540  
 QY 181 ProGlyValLysValValTyrLeuCyArgAsnProPheAspThrPheIleSerSerTyr 200  
 DB 541 CCGGATGTAAGGTCGTACTTGTGCGGAAACCGTTGACATTCATCTCTTCTGTG 600  
 QY 201 HisTyrThrAsnAsnIleLysSerGluSerValSerProValLeuLeuAspGlnAlaPhe 220  
 DB 601 CATTTACCAACCAACATCAATCCGATGAGTGAAGCCAGCTTGTGTAAGCAAGCTTTT 660  
 QY 221 AspLeuTyrCyArgGlyValIleGlyPheGlyProPheTyrGluHisMetLeuGlyTyr 240  
 DB 661 GATCTGATATGCGGGAGAGATCGGCTTGGCCGTTTGGGAGACATGTTGGAGTAC 720  
 QY 241 TTPArgGluSerLeuLysArgProGluLysValPhePheLeuArgTyrGluAspLeuLys 260  
 DB 721 TCGAGAGAGAGCTTGAAGAGACCAAGAAAGTCTTCTTTTAAGGTACGAGATCTCAA 780

QY 261 AspAspIleGlyThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuProPheThrGlu 280  
 DB 781 GACGACATGAGAGCAACTTGAAGAGGCTTGCACATTTCTTAGAGCTTCTTCAACGAA 840  
 QY 281 GluGluGluArgLysGlyValValLysAlaIleAlaGluLeuCySerPheGluAsnLeu 300  
 DB 841 GAAGAGCAAGAAAGGAGTGTGAAGGCTATGCCGACTGTGATGCTTCGAGAAATCTG 900  
 QY 301 LysLysLeuGluValLysLysSerAsnLysSerIleLysAsnPheGluLeuAspPheLeu 320  
 DB 901 AAGAACTTGAGGTGAACAACTCAACCAAGTCATCAAGAACTTGAAGATTCATTTCTT 960  
 QY 321 PheArgLysGlyValValSerAspTyrValAsnTyrLeuSerProSerGluValArg 340  
 DB 961 TTTCGAAAGAGAAAGTGAATGATGATTTGGCTTAATTTGTCACTTCAACAAGTGAAGA 1020  
 QY 341 LeuSerAlaLeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArgLeuSer 359  
 DB 1021 TTGTCAAGCTTAGTGAATGATGACCAAGTTAGGTGATCTGTCTCACTTCAAGTTGAGC 1077

RESULT 3  
 CQ832234 1273 bp DNA linear PAT 29-JUL-2004  
 LOCUS CQ832234  
 DEFINITION Sequence 1 from Patent WO2004057000.  
 ACCESSION CQ832234  
 VERSION CQ832234.1 GI:50831890  
 KEYWORDS  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE  
 AUTHORS Varin, L.  
 TITLE Methods and compositions for modulating tuber formation in  
 tuber-producing plants  
 JOURNAL Patent: WO 2004057000-A 1 08-JUL-2004;  
 Florisys Inc. (CA)

FEATURES  
 Source location/Qualifiers  
 1..1273  
 /organism="Arabidopsis thaliana"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:3702"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.72e-156 Length: 1273  
 Score: 1898.00 Matches: 359  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-10-019-931-3 (1-359) x CQ832234 (1-1273)

QY 1 MetAlaThSerSerMetLySerIleProMetAlaIleProSerPheSerMetCyHis 20  
 DB 66 ATGGTACCTCAAGCATGAAGACATTCATGCGATCCCAAGTTCTCATGTGTAC 125  
 QY 21 LysLeuGluLeuLeuLysGluGlyThrArgAspValProLysAlaGluGluAspGlu 40  
 DB 126 AAGCTGAGCTCTTAAAGAGCAAAACTCGGACGCTCCGAAAGCCGAAAGAGTGA 185  
 QY 41 GlyLeuSerCySgluPheGluGluMetLeuAspSerLeuProLysGluArgGlyTTPArg 60  
 DB 186 GGGCTAAGCTGCAGATTCCAAGAGATGTTGATTTCTTCTTAAGAGAGATGAGA 245  
 QY 61 ThrArgTyrLeuTyrLeuPheGluGlyPheTyrCySgluAlaLysGluIleGlnAlaIle 80  
 DB 246 ACTCGTACCTTACCTAATCCAAAGGTTTGGTGCCAAAGCAAGAGATTCAAGCATC 305  
 QY 81 MetSerPheGluLysHisPheGluInsLeuGluAsnAspValValLeuAlaThrIlePro 100

Db 306 ATGCTTTCACAAACATTTCCAAATCCCTCGAAGAACAGATCGTTCTCGCCACCATACCT 365  
Qy 101 LysSerGlyThrThrTTPLeuLysAlaLeuThrPheThrIleLeuAsnArgHisArgPhe 120  
Db 366 AAATCCCGGTACAACTGGCTTAAAGCTTTAACTTTCACATCCATTAACCGTCACCGGTTT 425  
Qy 121 AspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHisAspLeuVal 140  
Db 426 GATCCGGTTGCTCGAGTACCAACCACTCTTTTTCATCTCCAACTCTGATGACTTGTGA 485  
Qy 141 ProPhePheGlyTyrTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAla 160  
Db 486 CTTTCTTCGAGTCAAGCTTTACGCCAACGAGATGTTCCCGATCTCTCCGGTCTAGCC 545  
Qy 161 SerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGlyIuThrIleGlyLys 180  
Db 546 AGTCGAAGAACGTTTCGAAACCCACTTACCGTTCCGTTCCCTAAAGAAACATGAGAGAA 605  
Qy 181 ProGlyValLysValValTyrLeuCyArgAsnProPheAspThrPheIleSerSerTyr 200  
Db 606 CCCGGTGTGAAGTCTGTACTTGTGCGGAAACCGTTTGACATTCATCTTCCGTGG 665  
Qy 201 HisTyrThrAsnAsnIleLysSerGlySerValSerProValLeuLeuAspGlnAlaPhe 220  
Db 666 CATTAACACCAACATCAATCCGAGTCAGTGAACCCAGCTTGTAGACCAAGCTTTT 725  
Qy 221 AspLeuTyrCyArgGlyValIleGlyPheGlyProPheTyrGlnHisMetLeuGlyTyr 240  
Db 726 GATCTGATTTGCCCCGGAGATGATCGGTTTGCCCCGTTTGGAACACATGTGGAAATC 785  
Qy 241 TTPArgGlySerLeuLysArgProGlyLysValPhePheLeuArgTyrGlyLysAspLeuLys 260  
Db 786 TGGAGAGAGAGCTTGAAAGAGACAGAGAAAGTCTCTTTTAAAGTACAGAGATCTCAAA 845  
Qy 261 AspAspIleGlyThrAsnLeuLysArgLeuAlaThrPheLeuGlyLysProPheThrGly 280  
Db 846 GACACATCTCGAGACCACTTGAAAGAGGCTTGCAACTTCTTAGAGCTTCTTCAACCGAA 905  
Qy 281 GlnGlnGlyLysGlyValValIleAlaIleAlaGlyLeuLysSerPheGlyLysAsnLeu 300  
Db 906 GAAAGAGAACCAAGAGAGGAGTTGTGAAGGCTATCCCGAGCTGTGTAGCTTTCGAAGATCG 965  
Qy 301 LysLysLeuGlyValAsnLysSerAsnLysSerIleLysAsnPheGlyLysAsnArgPheLeu 320  
Db 966 AAGAGATTTGGAGTGGAACAAGTCAACAGTCAGATCAAGACTTTGAGAAATCGATTCCTG 1025  
Qy 321 PheArgLysGlyGlyValSerAspTyrValAsnTyrLeuSerProSerGlnValGlyArg 340  
Db 1026 TTTCGGAAGAGAGAGATGATGATGGGTTTAACTATTGTTCACCTTCACAAGTGGAAAGA 1085  
Qy 341 LeuSerAlaLeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArgLeuSer 359  
Db 1086 TTGTCACTCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1142

RESULT 4  
LOCUS CQ881446 1273 bp DNA linear PAT 11-OCT-2004  
DEFINITION Sequence 1 from Patent WO2004062350.  
ACCESSION CQ881446  
VERSION CQ881446.1 GI:54034405  
KEYWORDS  
SOURCE Arabidopsis thaliana (chale crese)  
ORGANISM Arabidopsis thaliana  
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; euroside II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE  
AUTHORS Maucher,H., Miersch,O., Masternack,C. and Varin,L.  
TITLE Methods and compositions for producing male sterile plants  
JOURNAL Patent: WO 2004062350-A 1 29-JUL-2004;  
Florisys Inc. (CA); Institut fuer Pflanzenbiochemie (DE)  
FEATURES  
^ source 1..1273

ORIGIN  
Alignment Scores:  
Pred. No.: 1,72e-156 Length: 1273  
Score: 1898.00 Matches: 359  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0  
US-10-019-931-3 (1-359) x CQ881446 (1-1273)

Qy 1 MetAlaThrSerMetLysSerIleProMetAlaIleProSerPheSerMetCyHis 20  
Db 66 ATGGCTAACCTCAAGCAAGAGATTCATGCGCATGCCAATGCCAAGTTTCTCCATGTGTAC 125  
Qy 21 LysLeuGlyLeuLeuLysGlyGlyLysThrArgAspValProLysAlaGlyLysAspGly 40  
Db 126 AAGCTCGAGCTCTTAAAGAGGCAAACTCGGAGCTCCCGAAAGCCGAAGAAAGAA 185  
Qy 41 GlyLeuSerCyGlyLysPheGlnGlyMetLeuAspSerLeuProLysGlyArgGlyTyrArg 60  
Db 186 GGGCTAAGTGGCGAGTTCAAGAGATGTGTGATTTCTCTCTTAAGAGAGAGATGGAGA 245  
Qy 61 ThrArgTyrLeuTyrLeuPheGlnGlyPheTyrCyGlnAlaLysGlyIleGlnAlaIle 80  
Db 246 ACTCGTTAATCTTAATCTTAATCAAGAGGTTTGGTCCCAAGCCAAAGAGATTCAAGCATC 305  
Qy 81 MetSerPheGlnLysHisPheGlnSerLeuGlyLysAsnAspValValLeuAlaThrIlePro 100  
Db 306 ATGTCTTTCACAAACATTTCCAAATCCCTCGAAACGACATCTGTTCCGCCACCATACCT 365  
Qy 101 LysSerGlyThrThrTTPLeuLysAlaLeuThrPheThrIleLeuAsnArgHisArgPhe 120  
Db 366 AAATCCCGGTACCACTGGCTTAAAGAGCTTTAACTTTCACATCCCTTAACCGTCAACCGGTTT 425  
Qy 121 AspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHisAspLeuVal 140  
Db 426 GATCCGGTTGCTCGAGTACCAACCACTCTTTCATCTTTCACATCCCTCAATGACTGTGA 485  
Qy 141 ProPhePheGlyTyrTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAla 160  
Db 486 CTTTCTTCGAGTACAGACTTTAGCCCAACGAGATGTTCCGATCTCTCGGGTCTAGCC 545  
Qy 161 SerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGlyIuThrIleGlyLys 180  
Db 546 AGTCGAAGAACGTTTCGAAACCCACTTACCGTTCCGTTCCCTAAAGAAACGATCGAAGAA 605  
Qy 181 ProGlyValLysValValTyrLeuCyArgAsnProPheAspThrPheIleSerSerTyr 200  
Db 606 CCCGGTGTGAAGTCTGTACTTGTGCGGAAACCGTTTGAACACATTCATCTTCCGTGG 665  
Qy 201 HisTyrThrAsnAsnIleLysSerGlySerValSerProValLeuLeuAspGlnAlaPhe 220  
Db 666 CATTAACCAACCAACATCAATCCGAGTCACTGAGCCCATCTTGTCAAGCAAGCTTTT 725  
Qy 221 AspLeuTyrCyArgGlyValIleGlyPheGlyProPheTyrGlnHisMetLeuGlyTyr 240  
Db 726 GATCTGATTTGCCCCGGAGATGATCGGTTTGCCCCGTTTGGGAACACATGTTGGGATAC 785  
Qy 241 TTPArgGlySerLeuLysArgProGlyLysValPhePheLeuArgTyrGlyLysAspLeuLys 260  
Db 786 TGGAGAGAGAGCTTGAAAGAGACAGAGAAAGTCTCTTTTAAAGTACGAGATCTCAAA 845  
Qy 261 AspAspIleGlyThrAsnLeuLysArgLeuAlaThrPheLeuGlyLysProPheThrGly 280  
Db 846 GACGACATCGAGACCACTTGAAAGAGCTTGCAACTTTTAAAGATCTTTCACCGAA 905  
Qy 281 GlnGlnGlyLysGlyValValIleAlaIleAlaGlyLeuLysSerPheGlyLysAsnLeu 300

Dh 906 GAAGAGGAAGAAAGGAGTTGTGAAGCTATCGCCAGCTGTGTAGCTTCGAGATCTG 965  
Qy 301 LysylalauGLuValAsnLysSerAsnLysSerIleLysAsnPhelGluAsnArgPheLeu 320  
Db 966 AAGAGGTGGAGTGAACCACTCAACAGACGATCAAGAACTTTGAGAAATCGATTCCTG 1025  
Qy 321 PheArgLysGlyGluValAsnSerAspTyrValAsnTyrLeuSerProSerGlnValGluArg 340  
Db 1026 TTTCGAAAGAGAAAGTGAAGTGGATTGGCTTACTTATTTGTCACTTCACAGTGAAGA 1085  
Qy 341 LeuSerAlaLeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArgLeuSer 359  
Db 1086 TTGTCAAGCTTAGTGAATGACAGATTAGTGTGATCTGCTCACCTTCAGGTTGAGC 1142  
RESULT 5  
LOCUS AY084999 1273 bp mRNA linear PLN 14-APR-2003  
DEFINITION Arabidopsis thaliana clone 124067 mRNA, complete sequence.  
ACCESSION AY084999  
VERSION AY084999.1 GI:21403709  
KEYWORDS  
SOURCE  
ORGANISM  
Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
1 (bases 1 to 1273)  
Haas,B.J., Volkovskiy,N., Town,C.D., Troughan,M., Alexandrov,N.,  
Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.  
Full-length messenger RNA sequences greatly improve genome  
annotation  
Genome Biol. 3 (6), RESEARCH0029 (2002)  
12093376  
JOURNAL PUBMED  
2 (bases 1 to 1273)  
REFERENCE Brover,V., Troughan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and  
AUTHORS Feldmann,K.  
TITLE Full-length cDNA from Arabidopsis thaliana  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 1273)  
AUTHORS Brover,V., Troughan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and  
Feldmann,K.  
TITLE Direct Submission  
JOURNAL Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,  
COMMENT Malibu, CA 90265, USA  
This clone sequence is one of 5,000 Ceres full-length cDNAs made  
available to TIGR and GenBank. The following quality assessment of  
this set was done by comparison with known proteins: two percent of  
the clones are estimated to be 5'-truncated; less than one percent  
are 3'-truncated; approximately two percent represent alternative  
splice variants, including unspliced introns and spliced exons; one  
percent may contain premature stop codons; five percent may have  
frame shifts in a coding region. A sequence is considered to be  
5'-truncated if it lacks the translation initiation start (ATG). A  
sequence is considered to be 3'-truncated if it lacks the  
C-terminal end of the encoded protein. Please note that these cDNA  
sequences are derived from the W8 or W8er ecotypes and therefore  
may contain polymorphisms when compared to sequences from Col-0.  
GenSet carried out the library production and sequencing of the  
full-length clones. Ceres, Inc. carried out the clustering of the  
5' sequences, selection of clones, and sequence assembly.  
FEATURES  
source  
location/Qualifiers  
1..1273  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/clone="124067"  
66..1145  
/codon\_start=1  
/product="steroid sulfotransferase-like protein"  
/protein\_id="AA061557.1"  
/db\_xref="GI:21537216"  
/translation="MATSMKSIEMAIIPFEMCHKLDELKEKTRDVPKAEDEGLSC  
EFGEMDSLPKERGWRTRYLYLFGQFMCQAKELQAIMSPQHGFSLENDVVLATIPRS

ORIGIN

Alignment Scores:  
Pred. No.: 1,72e-156 Length: 1273  
Score: 1896.00 Matches: 359  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 15 Gaps: 0

US-10-019-931-3 (1-359) x AY084999 (1-1273)

Qy 1 MetAlaThrSerSerMetLysSerIleProMetAlaIleProSerPheSerMetCysHis 20  
Db 66 ATGGCGACCTCAAGCAAGCAAGAGCATTCAGATGGCCATCCCAAGTTCTCTCATGTGTAC 125  
Qy 21 LysLeuGluLeuLeuLysGlyLysThrArgAspValProLysAlaGluGluAspGlu 40  
Db 126 AAGCTCAGACTCCTTAAAGAGCAAACTCGCAGCTCCCGAAGCCGAAGAGATGAA 185  
Qy 41 GlyLeuSerCysGluPheGlnGluMetLeuAspSerLeuProLysGluArgGlyTyrArg 60  
Db 186 GGGCTAAGCTCGCGATTCGAAGATTTGATTTCTCTTAAGAGAGAGATGAGA 245  
Qy 61 ThrArgTyrLeuTyrLeuPheGlnGlyPheTyrCysGlnAlaLysGluIleGlnAlaIle 80  
Db 246 ACTGTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTAC 305  
Qy 81 MetSerPheGlnLysHisPheGlnSerLeuGluAsnAspValLeuAlaThrIlePro 100  
Db 306 ATGCTTCCAAAACATTTCCAAATCCCTCGAAGACACGCTCTTCGCGACCATACCT 365  
Qy 101 LysSerGlyThrThrTyrPheLysAlaLeuThrPheThrIleLeuAsnArgHisArgPhe 120  
Db 366 AATATCCGTCACACCTGCTTAAAGCTTAACTTAACTTAACTTAACTTAACTTAACTTAA 425  
Qy 121 AspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHisAspLeuVal 140  
Db 426 GATCGGTTGCTCGATGATCAACCAACCTCTTTCATCTTCAACCTCAAGTGA 485  
Qy 141 ProPhePheGluTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAla 160  
Db 486 CTTTCTTCGAGTCAAGACTTACGCAAGAGATTTCCGATCTCTCGGCTTACGCC 545  
Qy 161 SerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGluThrIleGluLys 180  
Db 546 AGTCAGAAAGCTTGGACCACTTACCTTCCGTTCTTAAAGAAACATGAGAA 605  
Qy 181 ProGlyValLysValValTyrLeuCysArgAsnProPheAspThrPheIleSerSerTyr 200  
Db 606 CCCGTTGAAAGTCGTGACTTGTGCGGAACCTTTGACATTCATTCATTCGCG 665  
Qy 201 HisTyrThrAsnAsnIleLysSerGluSerValSerProValLeuLeuAspGlnAlaPhe 220  
Db 666 CATTAACCAACCAACATCAATCCGATCGACGACCTTGTGACCAAGCTTTT 725  
Qy 221 AspLeuTyrCysArgGlyValIleGlyPheGlyProPheTyrPgiLysIleGlyTyr 240  
Db 726 GATCTGATTTGCGCGGAGATGATGAGGTTTGGCCGTTTGGGAACATGTTGGATAC 785  
Qy 241 TrrArgGluSerLeuLysArgProGluLysValPhePheLeuArgTyrGluAspLeuLys 260  
Db 786 TGGAGAGAGAGCTTGAAGAGACCAAGAGTCTCTTTTAAAGTACGAGATCTCAA 845  
Qy 261 AspAspIleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuProPheThrGlu 280  
Db 846 GACGACATCAAGCAACATCTTGAAGAGCTTGAACATCTTGAAGCTTCTTCAACCGAA 905

QY		281	GlnGluGluAlaArglySerGlyValIleValIleValIleAlaGluLeuCySeSerPheGluAsnLeu	300
Dd		906	GAAGAAGAAACCAAAAGCGAGTGTGAAGGCATTCGCCGAGCTGTGTAGCTTGAGAGATTGG	965
QY		301	LysIleYseuGluaValAenIlySerSeraAnlysSerIleLysAanPheGluAsnArPheLeu	320
Dd		966	AAGAAGTTGACGCTGACCAACTCAAACAAGTCGATCAAGAACCTTTGAGAAATCGATTCTTG	1025
QY		321	PheArgLySerGlyGluValSerAspTrpValIeNtyrLeuSerProSerGlnValGluArg	340
Dd		1026	TTTTGGAAAGAGAGAAGTAGTGAATGGGTAACTTTGTCACCTTCCAAGTGGAAAGA	1085
QY		341	LeuSerAlaLeuValaAspAspIlySerGlyGlySerGlyLeuThrPheArgLeuSer	359
Dd		1086	TTGTCCAGCCTTAAGTGATGACCAAGTAACTGATCTGGTCTCACCTTCAGGTTGAGC	1142
RESULT 6				
LOCUS	AY099809	1347 bp	mRNA	linear
DEFINITION	Arabidopsis thaliana steroid sulfotransferase-like protein			
ACCESSION	(At5g07010) mRNA, complete cds.			
VERSION	AY099809			
KEYWORDS	AY099809..1 GI:20466685			
SOURCE	PLI CDNA.			
ORGANISM	Arabidopsis thaliana (chale creas)			
REFERENCE	Arabidopsis thaliana Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicotyledons; rosoids; euroside II; Brassicales; Brassicaceae; Arabidopsi.			
AUTHORS	Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H., Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A., Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K., Sakano,H., Sekurai,T., Satou,M., Seki,M., Shim,P., Yamada,K., Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.			
TITLE	Direct Submission			
JOURNAL	Submitted (24-APR-2002) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA			
COMMENT	e-mail for correspondence: arab@sequence.stanford.edu			
	RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sekurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.			
	The Salk, Stanford, PGSC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shim,P., Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.			
	Nguyen,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W. (SSP/Stanford) contributed equally to this work as PIs.			
FEATURES	location/qualifiers			
SOURCE	1..1347			
	/organism="Arabidopsis thaliana"			
	/mol_type="mRNA"			
	/db_xref="taxon:3702"			
	/chromosome="5"			
	/clone="RAF109-95-D03 (R25564)"			
	/ecotype="Columbia"			
	/note="This clone is in pBluescript"			
	1..1347			
	/gene="At5g07010"			
	66..1145			
	/gene="At5g07010"			
	/codon_start=1			
	/product="steroid sulfotransferase-like protein"			

DB	Query Match:	Score:	Percent Similarity:	Best Local Similarity:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
US-10-019-931-3 (1-359) x AY099809 (1-1347)		1.84e-156	1898.00	100.00%	1347	359	0	0	0	0
QY	1	MetaIatHseSerMetLysSerLlePromeAlaIleProSerPheSerMetCysHis			20					
Db	66	ATGCTACCTCCAAACATGAGAGCATTCCTCAATGGCGATCCCAAGTTCTCCATGTGTAC			125					
QY	21	LysLeuGluLeuLeuLysGluGlyLysThrArgAspValProLysAlaGluGluAspGlu			40					
Db	126	AAGCTCAAGCTCCCTTAAAGAGGCAAACTCGCACCTCCGAAAGCCGAAAGAGATGAA			185					
QY	41	GlyLeuSerCysGluPheGlnGluMetLeuAspSerLeuProLysGluLysGlyTyrArg			60					
Db	186	GAGGTAAAGCTGCGAGTTCCAGAGATTTGGATTCTTCTTAAAGAGAGATGAGAGA			245					
QY	61	ThrArgTyrLeuTyrLeuPheGlnGlyPheTyrCysGlnAlaLysGluIleGlnAlaIle			80					
Db	246	ACTCGTTACCTTATACCTATTCGAAGGTTTTGGTGGCAAGCAAGATTCGAAGCATC			305					
QY	81	MetSerPheGlnLysHisPheGlnSerLeuGluAsnAspValAlaLeuAlaThrIlePro			100					
Db	306	ATGCTTTCCAAAAACATTTCCATCCCTCGAATAACAGATCGTTCTTCGCCACATACCT			365					
QY	101	LysSerGlyThrThrTyrPheLysValAlaLeuThrPheThrIleLeuAsnArgHisArgPhe			120					
Db	366	AAATCCGATCAACCTGACTTAAAGCTTTAACTTTACCATCTTAACCGTCACCGGTTT			425					
QY	121	AspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHisAspLeuVal			140					
Db	426	GATCCGGTTGGCTGAGTACCAACCAACCTCTTTTCACTTCCAAACCTCATGACTTGTA			485					
QY	141	ProPhePheGluTyrTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAla			160					
Db	486	CCTTTCTTCGATGTCAGACTTTACGCCAAGGAGATTTCCCGATCTCTCGGCTGACC			545					
QY	161	SerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGluThrIleGluLys			180					
Db	546	AGTCAAGAAAGCTTGGCAACCACTTACCGTTCGGTTCCCTTAAAGAAAGATGAGAA			605					
QY	181	ProGlyValIlyAspValTyrLeuCysArgAsnProPheAspThrPheIleSerSerTyr			200					
Db	606	CCCGGTGGAAGCTGCTGACTTGTGCGGGAACCCGTTTGAACATTCATCTCTCGTGG			665					
QY	201	HisTyrThrAsnAsnIleLysSerGluSerValSerProValLeuLeuAspGlnAlaPhe			220					
Db	666	CATTACACCAACATCATCAATCCGAGTCAAGTACGACCACTCTGCTAAGACCAAGCTTTT			725					
QY	221	AspLeuTyrCysArgGlyValIleGlyPheGlyProPheTyrPheGluIleGlyTyr			240					
Db	726	GATCTGTAATGGCCGGGAGATGATGGGGTTTGGCCCGTTTGGGAACAATGTGGGATAC			785					
QY	241	TyrArgGluSerLeuLysArgProGluLysValPhePheLeuArgTyrGluAspLeuLys			260					
Db	786	TGGAGAGAGACTTGAAGAGACCAAGAAAGTCTTCTTTTAAAGTACGAGATCTCAAA			845					

Qy		261	Aap8ap1llec1utrrraenleuLyArqLeuaLamhrPhelenuGluLeuProphethGlu	280
Dd		846	GAGACACTCGAAGCAAACTGGAAAGAGGTTCGAACCTTCTTAGAGCTTCCATTCACCGAA	905
Qy		281	GUUGUjUGlurqlyrGSylVaIVaIlyeLaIIeLaGlueuCySerPheGIAsnleu	300
Dd		906	GAAGAGAACCAAGAGGAGTTGTGAAGCTATCCCGAGCTGTGTAGCTTGCAAAATCTG	965
Qy		301	LyeLaleuGIuValaenLYSeSeraNlysSerIlelySaenPheGIuAsnaYrPheIeu	320
Dd		966	AAGAATTGGAGGTGAACAAGTCAAAACAAGTCGATCAAGAAGACTTTGAGAAATCGATTCTTG	1025
Qy		321	PheaRglyeGIuValaSerASPTTPValaenTYleuSerPROSeGIuValaGIuAR	340
Dd		1026	TTCGGAAGAGAGAGTGAAGATTGGGTTTAACATTTGTCACCTTCACAAGTGGAAAGA	1085
Qy		341	LeuseRaIaleuValaApApelyLeuGIyGIyGerGIyleuTrrrPheaRlyeSer	359
Dd		1086	TTGTCAAGCTTAGTGCATGACAAAGTTAGTGCATCTGCTCATTTCAGGTTGAGC	1142
RESULT 7				
LOCUS	ABO10697/c	86380 bp	DNA	linear
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 5, pl clone:MOJ9.			
ACCESSION	ABO10697 BA000015			
VERSION	ABO10697.1 GI:2828182			
KEYWORDS				
SOURCE				
ORGANISM	Arabidopsis thaliana (thale cress)			
REFERENCE	Arabidopsis thaliana Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosoid II; Brassicales; Brassicaceae; Arabidopsie. 1 Kaneke,T., Kotani,H., Nakamura,Y., Sato,S., Asamizu,E., Miyajima,N. and Tabara,S. Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence features of the regions of 1,381,565 bp covered by twenty one physically assigned P1 and YAC clones DNA Res. 5 (2), 131-145 (1998) 9679202 2 (bases 1 to 86380) Nakamura,Y. Direct Submission Submitted (28-JUN-1998) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:yinakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934) Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MOJ9 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3/'), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremilni.zozi.lasstate.edu/cgi-bin/sp.cgi). Genes encoding tRNAs are predicted by tRNAcan-se (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAcan-se/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MPH15 and the 3' clone is T2B14. Location/Qualifiers 1..86380 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /db_xref="taxon:3702" /chromosome="5"			
JOURNAL				
PUBMED				
AUTHORS				
COMMENT				
FEATURES				
SOURCE				

CDs

"/clone="MOJ9"  
/clone\_lib="Mitsui Pt"  
/ecotype="Columbia"  
join(1077, .1154,1263, .1322,1410, .1487,1579, .1624,  
1754, .2025,3170, .3424,5041, .5193)  
/note="gene id:MOJ9.1"  
/codon\_start=1  
/evidence=not\_experimental  
/product="b21 transcription factor-like protein"  
/protein\_id="BAB11142.1"  
/db\_xref="GI:9759540"  
/translation="MLMKGKNGKGLASPDHDIKSSDPPTLRRLAONREARKSRRLR  
KKAYVQRESRITLTLEOIORABQSGVFGGSLIGSQOOOAAVPMDEYRWLWEE  
QORLLNREAVATQGHLSNBLRMFVDTCLAHYHLINLKMAGTVPFHLISGAKVLP  
AERFLMWGFRPSIIKVINQLEPLEQOIVGI CGLQOSTDAEFAALSQGLEWKP  
SLSIYSVDSLPBPASAPLPRLSNFMGMSIALNKLALGAFVYLQADNLRHQTIRLN  
QLTTPROARCLLAAVEAFHRLQLALSLMLARPRDQ"  
complement(6176, .8560)  
/note="gene\_id:MOJ9.2"  
/codon\_start=1  
/evidence=not\_experimental  
/product="anthranilate phosphoribosyltransferase-like  
protein"  
/protein\_id="BAB11143.1"  
/db\_xref="GI:9759541"  
/translation="MAADGAKSODEYKLDKMPRELSERRPHGQRCGTGIGSERRA  
STYLNEAMFLYRVVAKDLPNPATSCDPDVEVKIGNYKGTGHEKRTINPENW  
QVFPFASKIKVQSTVEFVGRDKEWVTDEYIGKVPDMREVPDPSPKAPQWYLL  
EDRGESKKREWVVAWLGTDQADEAPDMWSDASVQGEVQSVRSKYVSPKMYL  
LRVVIVLEQDPEPSDRQPOPAFKVQOQOILKTKCPKNTPMNEDLVFAAER  
FEEQPLTVEKVTTPADDEVKGRILSPLSVPERKIDHRAVHSKMYLKEFGALBED  
KRHEIKESRSIRHLHLCLEGGVHWDESLTYSIDKLPARQIMSPITGLEVGLISADQ  
LSPKTKDGKATTPDVCVAKGQKQVNRTRII DSSSPKMEQYWEVYDCTVITLGV  
FDNCHLGSEKSNCAKYVDSRIKGVRI,RLSTLEADRIYTHSPYLVIQTGLKMGSEV  
QLAQRVETLSAHMIYLYGHPLPKMHYLPFTNQDSLRYQMSIVARLSAEP  
LRKXNVEMLDSDHMSMBSRKFRFVYSPFAGLIMSKWGLDVCYKNNPLTTFE  
HVLFFILICYELLPTPTLYMLPIGLMNRFRFRHNDHTYSALEASPELDDE  
FDTPEKSGQVVVFKRIDRLKSVAGRIQMWGDIATGERRQALSNRDPATLTPIY  
FCLVAAMSLYVLPKRIALAGMFMHMKPERSKSPASPFKRLPSKADCM"  
join(11349, .11869,11959, .12410)  
/note="gb|AA659827.1  
gene\_id:MOJ9.3"  
/codon\_start=1  
/evidence=not\_experimental  
/product="polygalacturonase inhibiting protein 1; Fgip1"  
/protein\_id="BAB11144.1"  
/db\_xref="GI:9759542"  
/translation="MDKATYCLFLFPLPTTCSKDLGNQNDKNTLLIKIKSLNNPY  
HNASDPDTCSSWYCLECGDATNNHVTALITISGQISQIPAEVDDLPILETLYVR  
KLNTNLTGITPTIAKLNLPIRLSNWLTGPIPDFTLSQLNLELSEFNDLSGSIPI  
SLSLTPIPLTELSELRNKLTSIPESGSPFGPIPDRLSHNQLSGSI PSLGINDPN  
RIDRNRNLDQASMLPGSKNTGWSIDSRMFOPLSKYDIPTLCLIDLNHNGIIGG  
NIPQWNTAQLQFNVNKLCGHPIPGKQLQFDSVSYFNHKLCLGAPLEICK"  
join(12917, .13457,13541, .13992)  
/note="gene\_id:MOJ9.4"  
/codon\_start=1  
/evidence=not\_experimental  
/product="polygalacturonase inhibiting protein"  
/protein\_id="BAB11145.1"  
/db\_xref="GI:9759543"  
/translation="MDKNTMLFLLSLTLLTSLAKDLCHKDKTLLIKIKSLNNPY  
HNASDPDTCSSWYCLECGDATNNHVTSLIIDGDISQIPAEVDDLPILETLYVR  
KLNTNLTGITPTIAKLNLPIRLSNWLTGPIPDFTLSQLNLELSEFNDLSGSIPI  
SLSLTPIPLTELSELRNKLTSIPESGSPFGPIPDRLSHNQLSGSI PSLGINDPN  
RIDRNRNLDQASMLPGKNTGWSIDSRMFOPLSKYDIPTLCLIDLNHNGIIGG  
NIPQWNTAQLQFNVNKLCGHPIPGKQLQFDSVSYFNHKLCLGAPLEICK"  
join(14719, .14919  
/note="unnamed protein product; gb|AA63169.1  
gene\_id:MOJ9.5  
similar to unknown protein"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="BAB11146.1"

CDs







Qy 301 LysylserleuGluValaIenIySerAenIySerIleIyAenPheGluAenArgPheIeu 320  
 Db 54115 AAGAAGTTGGAGGCGAACAGATCAAAAGTCGATCAAGAACTTGGAGAAATCGATTCTTG 54056  
 Qy 321 PheArgIyGlyGluValaIeSerAenIySerIleIyAenIySerProSerGluValaIeGluArg 340  
 Db 54055 TTTCGAGAAAGAGAGAGTGGATGATGGTTAACTTTTGTCACTTCACAAAGTGGAAAGA 53996  
 Qy 341 LeuSerAlaIeValaIaSpaApIySerIyGlySerGlyLeuThrPheArgLeuSer 359  
 Db 53995 TTTCAGAGCTTAAGTCGATGACAAAGTTAGTGATCTGTCTCACTTCAGGTGAGC 53939  
 RESULT 8  
 LOCUS AY054219 1270 bp mRNA linear PLN 30-SEP-2001  
 DEFINITION Arabidopsis thaliana AT5G07010/MOJ9\_18 mRNA, complete cde.  
 ACCESSION AY054219 GI:15809902  
 VERSION AY054219.1 GI:15809902  
 KEYWORDS FLI\_CDNA.  
 SOURCE Arabidopsis thaliana (chale cress)  
 ORGANISM Arabidopsis thaliana  
 Euaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 1270)  
 REFERENCE 1 (bases 1 to 1270)  
 AUTHORS Shinn, P., Chen, H., Cheuk, R., Kim, C.J., Koesema, E., Meyers, M.C.,  
 Bahh, J., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D.,  
 Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamiya, A.,  
 Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X.,  
 Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J.,  
 Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A.,  
 Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,  
 Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.  
 TITLE Arabidopsis cDNA clones  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1270)  
 AUTHORS Shinn, P., Chen, H., Cheuk, R., Kim, C.J., Koesema, E., Meyers, M.C.,  
 Bahh, J., Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D.,  
 Hayashizaki, Y., Ishida, J., Jiang, P.X., Jones, T., Kamiya, A.,  
 Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X.,  
 Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J.,  
 Pham, P.K., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A.,  
 Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,  
 Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-AUG-2001) Salk Institute Genomic Analysis Laboratory  
 (SIGNAL), Plant Biology Laboratory, The Salk Institute for  
 Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,  
 USA  
 COMMENT  
 RIKEN Genomic Sciences Center (GSC) members carried out the  
 collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN  
 Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J.,  
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,  
 Hayashizaki, Y. and Shinozaki, K.  
 The Salk, Stanford, PGSC (SSP) Consortium members carried out the  
 sequencing and annotation of the RAFL cDNAs: Shinn, P., Chen, H.,  
 Cheuk, R., Kim, C.J., Koesema, E., Meyers, M.C., Bahh, J., Bowser, L.,  
 Dale, J.M., Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G.,  
 Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M.,  
 Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Southwick, A.,  
 Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,  
 Davis, R.W., Theologis, A. and Ecker, J.R.  
 Shinn, P. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to  
 this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)  
 contributed equally to this work as PIs.  
 Location/Qualifiers  
 1..1270  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3702"

/chromosome="5"  
 /clone="RAF109-22-M04 (R13921)"  
 /ecotype="Columbia"  
 1..67  
 68..1147  
 /note="steroid sulfotransferase-like protein"  
 /codon\_start=1  
 /product="AT5G07010/MOJ9\_18"  
 /protein\_id="AA106879.1"  
 /db\_xref="GI:15809903"  
 /translation="MATSMSKSIIPMAIPSEFMCHKELEKGRDTPVKAEBEGSLC  
 EPOEMDLSPKRGMRTRVYLFGFGQAKELQAIWSFGKHFGSLNDVLAIPPS  
 GTTWLKLPTLIINRHRFDVASTNPLPTSNPDLPPEYKLYANGVPLDGLA  
 SPRTFAHLPRGSLKETIEKRGVAVVLPDPTFISNMHTNINIKSSGSPVLLD  
 AFDDYCRGVTFAPFPEHMLGYRESLKREKPFPLREBDKDIENLRKATFLD  
 PFTEERKGVVAIAELCSFENDKLEVNKSNKSNKINFNRFLEFRGEVSDWNYLS  
 PSQVERLSALVDDKLGSGSLFLRLS"  
 1148..1270  
 ORIGIN  
 3'UTR  
 Alignment Scores:  
 Pred. No.: 5,76e-156 Length: 1270  
 Score: 1892.00 Matches: 358  
 Percent Similarity: 99.72% Conservative: 0  
 Best Local Similarity: 99.72% Mismatches: 1  
 Query Match: 99.68% Indels: 0  
 DB: 15 Gaps: 0  
 US-10-019-931-3 (1-359) x AY054219 (1-1270)  
 Qy 1 MetAlaThrSerSerMetIySerIleProMetAlaIleProSerPheSerMetCysHis 20  
 Db 68 ATGGCTACCTCAACAGAGAGAGATTCACATGCGAGATCCCAAGTTCTCCATGTGTAC 127  
 Qy 21 LysleuGluLeuLeuLysGluGlyLysThrArgAspValProLysAlaGluGluAspGlu 40  
 Db 128 AAGCTCGAGCTCCTTAAGAAAGGCAAAATCGGACGTCGGAAGCGGAAGGAAGATGAA 187  
 Qy 41 GlyLeuSerCysGluPheGluGluMetLeuAspSerLeuPolyGluArgGlyTTPArg 60  
 Db 188 GGGCTAAGCTCGGAGTTCCAAAGATGTGGATTCCTCTTCTTAAGAGAGAGATGGGA 247  
 Qy 61 ThrArgTyrLeuTyrLeuPheGluGlyPheTyrCysGlnAlaLysGluIleAlaIle 80  
 Db 248 ACTGTTACCTTACCTTAATTCAGAGGTTTGTGTCGACCAAGCAAGATTCAGGCATC 307  
 Qy 81 MetSerPheGlnLysHisPheGlnSerLeuGluAenAspValLeuAlaThrIlePro 100  
 Db 308 ATGCTTTCCAAAACATTTTCATCCCTCGAANAACGCGTTCGCCACCATACCT 367  
 Qy 101 LysSerGlyThrThrTyrLeuLysAlaLeuThrPheThrIleLeuAenArgHisArgPhe 120  
 Db 368 AAATCCGCTCAACATCGGCTAAAGCTTAACTTACCAATCCCTTAACCGTCACCGGTTT 427  
 Qy 121 AspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHisAspLeuVal 140  
 Db 428 GATCCGGTGGCTGAGACCAACACCCTCTTTTCACTTCAACCTTCAGACCTTGTGA 487  
 Qy 141 ProPhePheGluTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAla 160  
 Db 488 CTTTCTTTCGAGTCAAGACTTACGCAACGAGATGTCCGATCTCTCGGGTCTACCC 547  
 Qy 161 SerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGluThrIleGluLys 180  
 Db 548 AGTCCAAAGACGTTCCGCAACCACTTACCGTTCGTTCCCTAAAGAAACATGAGAAA 607  
 Qy 181 ProGlyValLysValValTyrLeuCysArgAsnProPheAspThrPheIleSerSerTyr 200  
 Db 608 CCCGATGAAGGTCGTGATCTTGTGCGGAACCGTGTGACATTCATCTTCTGTGG 667  
 Qy 201 HisTyrThrAsnAlaIleLysSerGlySerValSerProValLeuLeuAspGlnAlaPhe 220  
 Db 668 CATTAACCAACCAATCAATCCGATGAGTGAAGCCAGTCTTGCTGACCAAGCTTTT 727

QY 221 AspleuTYrCYaRgGlyVallIeglyPheGlyProPheTrpGluHImetIleuGlyTyr 240  
Db 728 GATCGTATATGCGGAGAGTGAATCGAGTTGCCCGTTTGGAACACATGTGGGATAC 787  
QY 241 TTPATGtJuseLeuLYaRgProGluYValPhePheLeuAArgTyrGluAspLeuLYs 260  
Db 768 TCGAGAGAGAGCTTGAAGAGCAGAGAAAGCTTCTTTTAAAGTACGAGATCTCAAA 847  
QY 261 AspAspIleGluThrAsnLeuLYaRgLeuValaThrPheLeuGluLeuProPheTrpGlu 280  
Db 848 GACGACATCGAACCAACTTGAAGAGGCTTGCACATTTCTTGAAGCTTCTTCAACCAA 907  
QY 281 GluGluGluYValGlyGlyVAlValLYaValIleValGluLeuCYaSerPheGluAsnLeu 300  
Db 908 GAAAGGAAACGAAAGAGAGTGTGAAGGCTATGCGAGCTGTGATGCTTCAGAAATCTG 967  
QY 301 LYsLYsLeuGluValaLYaLYsSerAsnLYsSerIleLYsAsnPheGluAsnAArgPheLeu 320  
Db 968 AAGAAGTTGGAGGTGAACAAGTCAACAAAGTCAATCAAGAACTTTGAGATGATTTCTTG 1027  
QY 331 PheATgLYGlyGlyValaLYsAspTrpValaLeuTYrLeuSerProSerGlnValGluArg 340  
Db 1028 TTTCCGAAAGGAGAGTGAATGTGGGTTAACTATTGTCACTTCAACAGTGAAGAA 1087  
QY 341 LeuSerAlaLeuValaAspAspLYsLeuGlyGlyLYsSerGlyLeuThrPheArgLeuSer 359  
Db 1088 TTGTACGCTTATGTGATGAACAAAGTTAAGTGAATCTGTCTCACTTTCAGTTGAGC 1144

RESULT 9 AKI17202 1236 bp mRNA linear PLN 14-FEB-2004  
LOCUS AKI17202  
DEFINITION Arabidopsis thaliana At5g07000 mRNA for putative steroid  
sulfotransferase, complete cds, clone: RAFL6-75-F02.

ACCESSION AKI17202.1 GI:26449503  
VERSION FLI CDNA; CAP trapper.  
KEYWORDS Arabidopsis thaliana (thale cress)  
SOURCE Arabidopsis thaliana  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1  
Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Iehida, J.,  
Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P.,  
Kawai, Y., Hayashizaki, Y. and Shinzaki, K.  
Arabidopsis thaliana full-length cDNA  
Published Only in Database (2002)

2 (bases 1 to 1236)  
Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Iehida, J.,  
Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P.,  
Kawai, Y., Hayashizaki, Y. and Shinzaki, K.  
Direct Submission  
Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences  
Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa  
230-0045, Japan (E-mail: msekigsc.riken.go.jp,  
URL: http://pifweb.gsc.riken.go.jp, Tel: 81-45-503-9625,  
Fax: 81-45-503-9586)

COMMENT An Arabidopsis full-length cDNA library was constructed essentially  
as reported previously (Seki et al. (1998) Plant J. 15:707-720,  
Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI  
and XhoI was ligated to modified lambda PUC-1-E vector (Carninci et  
al. (2001) Genomics 77:79-90) digested with BamHI and SalI.  
This clone is in a modified pBluescript vector.  
Please visit our web site (http://pifweb.gsc.riken.go.jp/) for  
further details.

FEATURES  
source location/Qualifiers

1..1236  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/chromosome="5"  
/clone="RAFL6-75-F02"

gene  
CDS  
/ecotype="Columbia"  
/note="common name: thale cress"  
1..1236  
/gene="At5g07000/MOJ9\_17"  
66..1109  
/gene="At5g07000/MOJ9\_17"  
/product="putative steroid sulfotransferase"  
/protein\_id="PAC1878.1"  
/db\_xref="GI:26449504"  
/translation="MAIPSPSMCKPPLLBKSGEGEBEGSLSTEFQMDLSLPKENG  
RRNNVLYLFGFRQAREIQATISFOGFGSLPDDVLATIPKSGITWMLATITLITL  
RHRDPVSSSDHPDLTNSNDHDLVPEPEKYLNGNVPDLGSLSPRTPTHTVPGA  
LKQSDVNSVAVVLYCNPPTFTSMWHYINNTSSEAVLDEAFDLCRGLIGF  
GPFEHNLGYWRSLEKRPKYLPLKVEDLKEDITETNKKLASPLGLPFTBEERQGVV  
KATIDCSFENLKLELVNKSskLIQNTENRPLFKSGVSDLVNLTFSQVERLSALVD  
DKLAGSGTLPRLS"

## ORIGIN

## Alignment Scores:

Pred. No.: 7,36e-129 Length: 1236  
Score: 1582.00 Matches: 305  
Percent Similarity: 92.46% Conservative: 26  
Best Local Similarity: 85.20% Mismatches: 21  
Query Match: 83.35% Indels: 6  
DB: 15 Gaps: 3

US-10-019-931-3 (1-359) x AKI17202 (1-1236)

QY 4 SerSerMetLYsSerIleProMetAlaIleProSerPheSerMetCYaHImetLeuGlu 23  
Db 45 TCAAGATCAAAAGACATTTCCAAATGCGATCCCAAGTTTCTCATGTGTCAACAAAGCCG 104  
QY 24 LeuLeuLYsGlyGlyYrThrArgAspValProLYaValaGluGluAspGlyLeuSer 43  
Db 105 CTCCCTTAAGGAGGCAAAAGC-----GAAAGCCAAAGAAAGAAAGGAGCTAAAGC 152  
QY 44 CYsGluPheGluGluMetLeuAspSerLeuProLYsGluArgGlyTTPArgThrArgTyr 63  
Db 153 TACAGATTCCAAAGAGATTTGAGATCTTCTCTTAAGAGAGAGAGACGAGAAATCGTTAC 212  
QY 64 LeuTYrLeuPheGluGlyPheTrpCYsGlnAlaLYsGluIleGlnAlaIleMetSerPhe 83  
Db 213 CTTTACTTATTCCAAAGGATTTCCGTGCGCAAGCTTAAGAGATTCAGCTATACCTCTTTC 272  
QY 84 GlnLYsHISpHeGlnSerLeuGluAsnAspValaValaLeuAlaThrIleProLYsSerGly 103  
Db 273 CAAAAACATTTTCAGTCCCTTCCAGACGACGATGTCTCGCACCATACCTAAATCTGGC 332  
QY 104 ThrTrpTrpLeuLYsAlaLeuThrPheThrIleLeuAsnArgHISArgPheAspProVal 123  
Db 333 ACAACCTGTGTTAAAGCTTTAACTTTACACATCTCTTACCGGTCAATCGGTTTGATCCGTT 392  
QY 124 ---AlaSerSerThrAsnHISProLeuPheThrSerAsnProHISAspLeuValProPhe 142  
Db 393 TCCCATCAATGATTCGACCAACCTCTTCTCAATCAACCACTTCGACGACCTGTGATCTTTC 452  
QY 143 PheGluTYrLYsLeuTYrAlaAsnGlyAspValaProAspLeuSerGlyLeuAlaSerPro 162  
Db 453 TTCGAGTACAAAGCTTACGCGCAAGGAAATGTTCCGATTCCTGGAGCTACGAGTCCA 512  
QY 163 ArgThrPheAlaThrHISLeuProPheGlySerLeuLYsGluThrIleGluLYsProGly 182  
Db 513 AGAACATTCGCAACCAACGATACGTTGCGTCCCTTAAAGATTCGATGAGAAATCCCACT 572  
QY 183 ValLYsValaLYrLeuCYaArgAsnProPheAspThrPheIleSerSerTrpHisTYr 202  
Db 573 GTGAAGGTGTGTACTGTGCGGAAACCGTTTACACATTCATCTCATGTGGATGATAC 632  
QY 203 ThrAsnAsnIleLYsSerGluSerValSerProValleuAspGlnAlaPheAspLeu 222  
Db 633 ATCAACCAACATCACTTCGAGATCGATGAGCGCAAGCTTGTGACGAAGCTTTTGATCTA 692

Qy 223 TrrCysArpGly---ValIleGlyPheGlyProPheTrpGluHsIsmEtleuGlyTyrTrp 241  
Db 693 TATTGGCGGGATTTACTATCGATTTGGCCGTTTGGAAACATGTTGGATTCTGG 752  
Qy 242 ArgIleSerLeuLysArgProGluLysValPhePheLeuArgTyrGluAspLeuLysAsp 261  
Db 753 AGAGAGAGCTTGAAAGAGCCAGAAAGTCTTATTTTAAAGTAAGAGATCTCAAGAA 812  
Qy 262 AspIleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuProPheThrGluGlu 281  
Db 813 GACATCGAGCAACCTTGAAAGAGCTTACCAAGTCTTACCAAGTCTTACCAAGAA 872  
Qy 282 GluGluArgLysGlyValValIleValIleAlaIleAlaGluLeuCysSerPheGluAsnLeuLys 301  
Db 873 GAGGACAAAGAGGAGTTGTGAAGCTATCGCTGATGCTGTGAGCTTTGAGAAATCTGAAG 932  
Qy 302 LysLeuGluValAsnLysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeuPhe 321  
Db 933 AAGTTGGAGCTGAACAAGTCAAGCAAAATTGATCCAGACTATGAGAACCGGTTCTGTTT 992  
Qy 322 ArgLysGlyGluValSerAspTrpValAsnTyrLeuSerProSerGluValGluArgLeu 341  
Db 993 AGGAAAGAGAAAGTGAAGTATTTGGTTAACTATTTGTGCGCATCAAGAGTGAAGATG 1052  
Qy 342 SerAlaLeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArgLeuSer 359  
Db 1053 TCAGCTTGAAGTGAAGACAAAGTACGCTGATCTGCTCAGCTTCAAGATTAGT 1106  
RESULT 10  
LOCUS AX080460 1041 bp DNA linear PAT 22-FEB-2001  
DEFINITION Sequence 3 from Patent WO0102589.  
ACCESSION AX080460  
VERSION AX080460.1 GI:13159886  
KEYWORDS  
SOURCE Arabidopsis thaliana (chale crese)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophytes; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE 1  
AUTHORS Varin,J. and Gidda,S.  
TITLE Methods, compositions and genetic sequences for modulating  
flowering in plants, and plants genetically modified to flower  
early and tardily  
JOURNAL Patent: WO 0102589-A 3 11-JAN-2001;  
Varin, Luc (CA) ; Gidda, Satinder (CA)  
FEATURES  
source  
1. 1041  
/organism="Arabidopsis thaliana"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:3702"  
ORIGIN  
Alignment Scores:  
Score: 2.08e-126 Length: 1041  
1553.00 Matches: 299  
Percent Similarity: 92.31% Conservative: 25  
Best Local Similarity: 85.19% Mismatches: 21  
Query Match: 81.82% Indels: 6  
DB: Gaps: 3  
US-10-019-931-3 (1-359) x AX080460 (1-1041)

Db 109 GACTCTCTTCCCTAAGAGAGAGAGCGAGAAATCGTTACCTTATCTTATCCAAAGGTTT 168  
Qy 71 TrpCysGlnAlaLysGluIleGlnAlaIleIleSerPheGlnLysHsIsmPheGlnSerLeu 90  
Db 169 CGGTGCCAAGCTTAAGAGATTCAAGCTATCAAGCTTTCCAAAAAATTTTCACTCCCTT 228  
Qy 91 GluAsnAspValValIleuAlaThrIleProLysSerGlyThrTrpLeuValAlaLeu 110  
Db 229 CCAGACGACGTTGCTCGCCACATACCTTAATCTGACACACCTGTTAAAGCTTTA 288  
Qy 111 ThrPheThrIleLeuAsnArgHsIsmArgPheAspProVal---AlaSerSerThrAsnHis 129  
Db 289 ACTTTCACACATCTTACCGGTATCGGTATGATCGGTTTCCATCAAGTTCGAGCAC 348  
Qy 130 ProLeuPheThrSerAsnProHisAspLeuValProPhePheGluTyrLysLeuTyrAla 149  
Db 349 CCTCTTCAACATCAACCCCTCAAGACCTCGTACCTTCTTCCGAGTACAAAGCTTTACGCC 408  
Qy 150 AsnGlyAspValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeu 169  
Db 409 AACGAAATGTTCCGATCTCTCGGGTCTAGCCAGTCCAGAAACATTCGCAACCCAGTA 468  
Qy 170 ProPheGlySerLeuLysGluThrIleGlyLysProGlyValValIleTyrLeuCys 189  
Db 469 CCGTTCGGTGGCTTTAAGATTCGGTGCAGAAATCCAGTGGAAAGTTGTACTGTGTC 528  
Qy 190 ArgAsnProPheAspThrPheIleSerSerTrpHisTyrThrAsnAsnIleLysSerGlu 209  
Db 529 CGGAACCCGTTTGACATTCATCTCATGCGCATTAACATCAACAAATCATCTCCGAG 588  
Qy 210 SerValSerProValLeuLeuAspGlnAlaPheAspLeuTyrCysArgGly---ValIle 228  
Db 589 TCAGTGAAGCGCAGCTTGTGACGAAGCTTTGATCATTAATTCGCGGAGTTACTGTATC 648  
Qy 229 GlyPheGlyProPheTrpGluHsIsmEtleuGlyTyrTrpArgIleSerLeuLysArgPro 248  
Db 649 GGAATTGGCCCGCTTTTGGAAACATGTTGGGATACAGAGAGAGAGTTCGAAGCGCA 708  
Qy 249 GluLysValPhePheLeuArgTyrGluAspLeuLysAspIleGluThrAsnLeuLys 268  
Db 709 GAGAAAGCTTATTTTAAAGTACGAGATCTCAAAAGACATCGAGACCACTTGAAG 768  
Qy 269 ArgLeuAlaThrPheLeuGluLeuProPheThrGluGluGluValValIle 288  
Db 769 AAGCTACGAGTCTTTCAGACATCTTTCACCGAAGAGAGAAACAAAGAGAGTTGTG 828  
Qy 289 LysAlaIleAlaGluLeuCysSerPheGluAsnLeuLysLysLeuGluValAsnLysSer 308  
Db 829 AAAAGCTATCGCTGATCTGTGAGCTTTGAGAAATCTGAAGAGTTGAGAGTGAACAAGTCA 888  
Qy 309 AsnLysSerIleLysAsnPheGluAsnArgPheLeuPheArgLysGlyValValSerAsp 328  
Db 889 AGCAAAATTGATCCAGAACTATGAGAACCGGTTCTTGTGGAAGAGGAGAGTGAATGAT 948  
Qy 329 TrpValAsnTyrLeuSerProSerGluValGluArgLeuSerAlaLeuValAspAspLys 348  
Db 949 TTGGTTACATATTTGTGCGCATCAAGTGAAGAGATTGTGAGCTTGAAGTGAACAAG 1008  
Qy 349 LeuGlyLysSerGlyLeuThrPheArgLeuSer 359  
Db 1009 TTAGCTGATCTGCTCAGCTTTCAGATTGAGT 1041  
RESULT 11  
LOCUS BT006241 1044 bp mRNA linear PLN 17-APR-2003  
DEFINITION Arabidopsis thaliana At5g07000 mRNA, complete cds.  
ACCESSION BT006241  
VERSION BT006241.1 GI:30017312  
KEYWORDS FLI cDNA.  
SOURCE Arabidopsis thaliana (chale crese)  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE AUTHORS	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurooids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 1044) Kim,C.J., Chen,H., Cheuk,R., Shim,P., Bowser,L., Carninci,P., Chen,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Heun,V.W., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onders,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.
TITLE JOURNAL REFERENCE AUTHORS	Arabidopsis ORF clones Unpublished 2 (bases 1 to 1044) Kim,C.J., Chen,H., Cheuk,R., Shim,P., Bowser,L., Carninci,P., Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Heun,V.W., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Onders,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.
TITLE JOURNAL	Direct Submision Submitted (17-Apr-2003) Salk Institute Genomic Analysis Laboratory (Signal), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
COMMENT	RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN Arabidopsis Full-Length cDNA' : Seki,M., Narusaka,M., Ishida,J., Hayashizaki,Y. and Shinozaki,K.  The Salk, Stanford, RGC (SSP) Consortium members constructed and sequenced the PUN1 (ORF) clones using the RAPL cDNAs: Kim,C.J., Chen,H., Cheuk,R., Shim,P., Bowser,L., Chan,M.M., Chang,C.M., Dale,J.M., Heun,V.W., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onders,C.S., Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K., Yu,G., Yuan,S., Davis,R.W., Theologis,A., and Ecker,J.R.
FEATURES source	Kim,C.J. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk) contributed equally to this work as PIs.  location/Qualifiers 1..1044 /organism="Arabidopsis thaliana" /mol_type="mRNA" /db_xref="taxon:3702" /chromosome="5" /clone="U60156" /ecotype="Columbia" /note="This clone is in PUN1 51" 1..1044 /note="putative steroid sulfotransferase" /codon_start=1 /product="At5g07000" /protein_id="AAPI2890.1" /db_xref="GI:30017313" /translation="MAIPSFMCNRPPELLKEGSGQEEGLSYEFQEMDLSLPKRGRRNRRLYFOGRCQAKIQAITSFOKFIQIPDVVLAITPKSSTMLKALFTILLRHRPVPSSSDHPLTNSNPDLVFPFESYLKANGVPDLSGASPTPATYVFGALKDSVNSVYVYICRNPDPFTISMMHYINNTISVSAYLVDAAPLYCGRLIGFGPFWHMIGYRESLKREKVIPLKTELDKEDIEFNKKLASFLGLPTEEBQKGVKAIADLCFENDKLKVNKSSKLIQNYENRFLFRGEVSDLVNYSFQVERLSLVD DKLASGLTFRLS"
ORIGIN	
Alignment Scores:	
Pred. No.:	2,09e-126
Score:	1553.00
Percent Similarity:	92.31%
	Length: 1044 Matches: 299 Conservative: 25

Best Local Similarity:	85.19%	Mismatches:	21
Query Match:	81.82%	Indels:	6
DB:	15	Gaps:	3
US-10-019-931-3 (1-359) x BF006241 (1-1044)			
QY	11	MeCAlaIleProSerPheSerMetCysH1slyLeuGluLeuLyglGlyLeuThr	30
DB	1	ATGGGAGATCCCAAGTTTCTCATGTGTACAAAGCCGAGCTCTTAAAGAAAGCAAGC	60
QY	31	ArgAspValProLyValGluGluAspGluGlyLeuSerCysgluPheGluMetLeu	50
DB	61	-----GAAGGCCAAGAAAGAAAGGCTTAAGCTACGATTCGAAGATGTG	108
QY	51	AspSerLeuProLyValGluGlyValGluGlyValGluGlyValGluGlyValGluGly	70
DB	109	GACCTCTTCTTAAG	168
QY	71	TrpCysGlnAlaLyglGluGluGluGluGluGluGluGluGluGluGluGluGluGlu	90
DB	169	CGGAGCCAGCTAAGAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC	228
QY	91	GluAsnAspValValLeuAlaThrIleProLySerGlyThrThrProLyValAlaLeu	110
DB	229	CCAGACGACGTTGCTCGCACCATACCTAACTGGCACACCTGGTTAAAGCTTTA	288
QY	111	ThrPheThrIleLeuAsnArgHisArgPheAspProVal---AlaSerSerThrAsnHis	129
DB	289	ACTTCCACCATCTTACCGCATCGGTGATCGGTTCCTCATCAAGTTCCGACAC	348
QY	130	ProLeuPheThrSerAsnProHisAspLeuValProPhePheGluTrpValLeuTrpAla	149
DB	349	CCTCTTCCACATCCCAACCTCCACGACCTGTACTTCTTCCAGTCAAGCTTTACGCC	408
QY	150	AsnGlyAspValProAspLeuSerGlyLeuAlaSerProAlaGlyThrPheAlaThrHisLeu	169
DB	409	AACGAAATGTTCCGATCTCTCGGCTTACCGATCCAAAGAAATTCGCAACCCAGCTA	468
QY	170	ProPheGlySerLeuLyglGluThrIleGluLyProGlyValLyValValTrpLeuCys	189
DB	469	CCGTTCCGTCCTTAAAGATTCGATCGAAGATCCAGTGGAAAGTTGTACTCTGTC	528
QY	190	ArgAsnProPheAspThrPheIleSerSerTrpHisTrpThrAsnAsnIleLySerGlu	209
DB	529	CGGAACCGTTTGACATTCATCTCCATGTGCAATACATCAACATCATCTCCGAG	588
QY	210	SerValSerProValIleLeuAspGlnAlaPheAspLeuTrpCysArgGly---ValIle	228
DB	589	TCACTGAGCCGACGCTTGTGAGAGAGCTTTGATCTATATTCGCGGAGTTACTGATC	648
QY	229	GlyPheGlyProPheTrpGluHisMetLeuGlyTrpTrpArgGluSerLeuLyArgPro	248
DB	649	GGAATTTGCCCGTTTGGGAAACATGTTGGGATACCTGAGAGAGAGCTTGAAGGCCA	708
QY	249	GluLyValPhePheLeuArgTrpGluAspLeuLyAspAspIleGluThrAsnLeuLy	268
DB	709	GAGAAAGCTTATTTTAAAGTACGAGATCTCAAGAGACATCGAGACCAACTTGAAG	768
QY	269	ArgLeuAlaThrPheLeuGluLeuProPheThrGluGluGluValGlyValVal	288
DB	769	AAGCTAGCAAGTTCTTGAAGCTTTCACCAAGAGAGAGAAAGAGAGAGAGAGAGAG	828
QY	289	LyAlaIleAlaGluLeuCysSerPheGluAsnLeuLyLyLeuGluValAsnLySer	308
DB	829	AAAGCTATCGTGAATCTGTGAGCTTTGAGAAATCTGAAGAGTTGAGAGTGAACAAGCA	888
QY	309	AsnLySerIleLyAsnPheGluAsnArgPheLeuPheArgLyglGlyValSerAsp	328
DB	889	AGCAAAATGATCCAGAACTAGAGAACCGGTTCTTTAGAGAAAGAGAGAGAGATGAT	948
QY	329	TrpValAsnTrpLeuSerProSerGlnValGluArgLeuSerAlaLeuValAspAspLy	348
DB	949	TTGGTTAACTATTTGTGCGCATCACAAGTGAAAGATTTGACCTTAGTGATGACAAAG	1008

Qy 349 LeuGlyGlySerGlyLeuThrPheArgLeuSer 359  
Db 1009 TTAGCTGATCTGGTCTCATTTCAGATTGAGT 1041

RESULT 12  
CO8832235

LOCUS CO8832235 1023 bp DNA linear PAT 29-JUL-2004  
DEFINITION Sequence 2 from Patent WO2004057000.  
ACCESSION CO8832235  
VERSION CO8832235.1 GI:50831891  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1  
AUTHORS Varin,L.  
TITLE Methods and compositions for modulating tuber formation in  
JOURNAL tuber-producing plants  
PUBLISHER Patent: WO 2004057000-A 2 08-JUL-2004;  
Florisys Inc. (CA)  
FEATURES  
source Location/Qualifiers  
1..1023  
/organism="Arabidopsis thaliana"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:3702"

ORIGIN  
Alignment Scores:  
Pred. No.: 1.92e-123 Length: 1023  
Score: 1519.00 Matches: 292  
Percent Similarity: 92.15% Conservative: 25  
Best Local Similarity: 84.88% Mismatches: 21  
Query Match: 80.03% Indels: 6  
DB: Gaps: 3

US-10-019-931-3 (1-359) x CO8832235 (1-1023)

Qy 18 MetCysHisIysLeuGluLeuLeuGluGlySerThrArgAspValProIleAspGlu 37  
Db 1 ATGGTCAACAAGCCGAGCTCTTAAGGAAGCAAAAGC-----GAAGGCCAA 48

Qy 38 GluAspGluGlyLeuSerCysGluPheGlnIleuMetLeuAspSerLeuProIleGluArg 57  
Db 49 GAAGAAAGAGGGCTTAAGCTACGAGTCCAGAGATGTTGACTCTTCTTAAGAGAGA 108

Qy 58 G1TTPARGTHARGTYRLEUTYRLEUPHEGNGLYPHETPYSGINAlaIysGluIle 77  
Db 109 GGACGGAGAAATCTTAACCTTACTTATCCAGGGTTCCGTCCAGAGCTTAAGAGATT 168

Qy 78 GlnAlaIleMetSerPheGlnIysHisPheGlnSerLeuGluAsnAspValIleuAla 97  
Db 169 CAAGCTATACGCTTTCCAAAACATTTTCAGTCCCTCCAGAGAGCTTCTCTGCC 228

Qy 98 ThrIleProIysSerGlyThrThrTTPLeuIysAlaLeuThrPheThrIleLeuAsnArg 117  
Db 229 ACCATTAACCTTAATCTGGACAACTGTTAAACCTTAACCTTACCATCTTAACCGT 288

Qy 118 HisArgPheAspProVal---AlaSerSerThrAsnHisProLeuPheThrSerAsnPro 136  
Db 289 CATGGTTTGAATCCGGTTTCTCTCATCAAGTCCAGACCACTTCTTCACATCCAACT 348

Qy 137 HisAspLeuValProPhePheGluTyTrIysLeuTyTrAlaAsnGlyAspValProAspLeu 156  
Db 349 CAGACCTCGTAACCTTTCTTGAGATACAAGCTTTACGCCAAGAAATGTTCCCGATCTC 408

Qy 157 SerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPheGlySerLeuIysGlu 176  
Db 409 TCGGGTCTAAGCACTGCAAGAAATCTGCAACCAAGTACGTTCCGGTCCGCTTAAGGAT 468

Qy 177 ThrIleGluIysProGlyValIysValValTyTrIleuCysArgAsnProPheAspThrPhe 196

Db 469 TCGGTCAAGAAATCCCACTGTGAAGGTTGTGTACTGTGCGCGAAACCGTTTGACACTTC 528  
Qy 197 IleSerSerTrpHisTyTrThrAsnAsnIleIysSerGluSerValSerProValLeuLeu 216  
Db 529 ATCTCCATGTGGCATTCATCAACAACATCATCTCCAGTCCAGAGCGGACGCTTCTCTA 588

Qy 217 AspGlnAlaPheAspLeuTyTrCysArgGly---ValIleGlyPheGlyProPheTrpGlu 235  
Db 589 GACCAAGCTTTGATCTATATATGCGGGGATTAAGTATCGGATTTGGCCGTTTGGGAA 648

Qy 236 HisMetLeuGlyTyTrTPARGIusSerLeuIysPheArgProGluIysValPhePheLeuArg 255  
Db 649 CACATGTGGGATTAAGTGAAGAGAGCTTGAAGGCGCAAGAAAGCTTATTTTAAAG 708

Qy 256 TyTrGluAspLeuIysAspAspIleGluThrAsnLeuIysArgLeuAlaThrPheLeuGlu 275  
Db 709 TACGAGATCTCAAGAGAACATCGAGACCACTTGAAGAGCTACGACAGTTCTTGA 768

Qy 276 LeuProPheThrGluGluGluIysGlyValValIysAlaIleAlaGluLeuCys 295  
Db 769 CTTCCTTTCACCGAAGAGAGAAACAAAGGAGGATTTGAAAGCTATCGCTGATCTGTGT 828

Qy 296 SerPheGluAsnLeuIysIysLeuGluValAlaAsnIysSerAsnIysSerIleIysAsnPhe 315  
Db 829 AGCTTTAGAAATCTGAAGAAAGTTGAGCTGACACAGTCAAGCAAAATGATCCAGAACTAT 888

Qy 316 GluAsnArgPheLeuPheArgIysGlyGluValSerAspTrpValAsnTyTrLeuSerPro 335  
Db 889 GAGAACCGGTTCTTGTATTAAGAAAGAGAGAGTGAATGTTGATTATTTGTCCCA 948

Qy 336 SerGlnValIysArgLeuSerAlaLeuValAspAspIysLeuGlyIysSerGlyLeuThr 355  
Db 949 TCACAACTGGAAGAAATTTGACCTTATGATGATCAAGATTAGCTGATCTGATCTACT 1008

Qy 356 PheArgLeuSer 359  
Db 1009 TTCAGATTGAGT 1020

RESULT 13  
CO881447 1023 bp DNA linear PAT 11-OCT-2004  
LOCUS CO881447  
DEFINITION Sequence 2 from Patent WO2004062350.  
ACCESSION CO881447  
VERSION CO881447.1 GI:54034406  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1  
AUTHORS Maucher,H., Miersch,O., Wasternack,C. and Varin,L.  
TITLE Methods and compositions for producing male sterile plants  
JOURNAL Patent: WO 2004062350-A 2 29-JUL-2004;  
Florisys Inc. (CA); Institut fuer Pflanzenbiochemie (DE)  
FEATURES  
source Location/Qualifiers  
1..1023  
/organism="Arabidopsis thaliana"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:3702"

ORIGIN  
Alignment Scores:  
Pred. No.: 1.92e-123 Length: 1023  
Score: 1519.00 Matches: 292  
Percent Similarity: 92.15% Conservative: 25  
Best Local Similarity: 84.88% Mismatches: 21  
Query Match: 80.03% Indels: 6  
DB: Gaps: 3

US-10-019-931-3 (1-359) x CO881447 (1-1023)

```

Oy 18 MetCyb5h1sLeuGluLeuLeuGluGlyThrArgAspValProlysa1aGlu 37
Db 1 ATGTGTCAAAAGCCGAGCTCTTAAAGAGCAAAAGC-----GAAGGCCAA 48
Oy 38 GluAepGluGlyLeuSerCyseGluPheGlnGluMetLeuAspSerLeuProlysa1aArg 57
Db 49 GAAGAAAGAGGAGCTTAAGTACGAGAGTTCAGAGAGAGTTCCTTCCAAAGAGAGA 108
Oy 58 GlyTPArgThrArgTyrluTyrluPheGlnGlyPheTPCyseGlnAlaGlyGluTle 77
Db 109 GAGCGAGAAATCGTTACCTTACTTATTCAGAGGTTCCGCTGCCAAGCTTAAGAGATT 168
Oy 78 GlnAlaIleMetSerPheGlnLysHisPheGlnSerLeuGluAenAspValValLeuAla 97
Db 169 CAAGCTATCACGCTTCCAAAAAACATTTTCAAGTCCCTTCAGAGAGAGAGCTGCTCCGC 228
Oy 98 ThrIleProlysaSerGlyThrThrTrtleuLysAlaLeuThrPheThrIleLeuAenArg 117
Db 229 ACCATACCTTAATCGGACAAACCTGGTTAAAGCTTTACATCCATCCTTAACCCGT 288
Oy 118 HisArgPheAspProVal---AlaSerSerThrAenHisProLeuPheThrSerAenPro 136
Db 289 CATCGGTTGATCCGGTTCCCTCATCAAGTTCCGACCAACCTCTTCTCAATCCAACTC 348
Oy 137 HisaPheLeuValProPhePheGluTyrluTyrluLysValaAenGlyAspValProAspLeu 156
Db 349 CACGACCTCGTACCTTCTCGAGTACAAAGCTTTCGCAACGAGAAAGTCCCGCATCTC 408
Oy 157 SerGlyLeuAlaSerProArgThrPheAlaThrHisleuProPheGlySerLeuLysGlu 176
Db 409 TCGGGTCTAAGCAGCCAGCAAGAACATTCGCAACCCGATCGGTGGTCCCTTAAGAT 468
Oy 177 ThrIleGluTyrluProGlyValLysValaValTyrluLeuCyAspAenProPheAspThrPhe 196
Db 469 TCGGTCGAGATCCCAAGTGAAGTTGTGTACCTGTGCTCCGGAACCTTTGACACATTC 528
Oy 197 IleSerSerTPHsTyrluThrAenAenLysSerGluSerValSerProValLeuLeu 216
Db 529 ATCTCCATGATGGCATTTATCATCAACAACATCTTCGAGTCAAGTGAAGCGCATCTTGCTA 588
Oy 217 AspGlnAlaPheAspLeuTyrluCyseArgGly---ValIleGlyPheGlyProPheTPGlu 235
Db 589 GACGAAAGCTTTGATCTATATTGTCGAGGATTAATCATGATCGATTGGCCGTTGGGAA 648
Oy 236 HisMetLeuGlyTyrluArgGluSerLeuLysArgProGluTyrluValPhePheLeuArg 255
Db 649 CACATGTTGGGATCTGAGAGAGAGCTTGAAGAGCCAGAGAAAGTCTTATTTTAAAG 708
Oy 256 TyrluAepLeuLysAspAspIleGluThrAenLeuLysArgLeuAlaThrPheLeuGlu 275
Db 709 TACGAGGATCTCAAGAAGACATCGAGACCAACTGAAGAAGCTAGCAAGTTTCTTAGA 768
Oy 276 LeuProPheThrGluGluGluLysArgLysValValLysAlaIleAlaGluLeuCyse 295
Db 769 CTTCCTTCCAGGAAGAGAGAACAAAGGAGAGTGTGAAGAGCTATGCTGATCTGCT 828
Oy 296 SerPheGluAenLeuLysLeuGluValAenLysSerAenLysSerIleLysAenPhe 315
Db 829 AGCTTTGAGATCTCAAGAAAGTGAAGGATGAACAAGTCAACGAATTTGATCAGAACTAT 888
Oy 316 GluAenArgPheLeuPheArgLysGlyLysValaSerAspTPPValAsnTyrluLeuSerPro 335
Db 889 GAGAAACCGGTTCTGTGTAGGAAGAGAGAGAGAGTGTGTTAACTATTTGTCGCCA 948
Oy 336 SerGlnValGluArgLeuSerAlaLeuValaAspAspLysLeuGlySerGlyLeuThr 355
Db 949 TCACAAGTGAAGAGATTGTCAAGCTTAGTGATGACAAAGTTAGCTGATCTGCTCACT 1008
Oy 356 PheArgLeuSer 359
Db 1009 TTCAAGATTGAGT 1020

```

```

AC135566
LOCUS      115071 bp      DNA      linear      HTG 13-AUG-2005
DEFINITION Medicago truncatula clone mth2-2298, WORKING DRAFT SEQUENCE, 4
unordered pieces.
ACCESSION AC135566.29 GI:72534892
VERSION    HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS   Medicago truncatula (Barrel medic)
SOURCE     Medicago truncatula
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
REFERENCE  1 (bases 1 to 115071)
AUTHORS    Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE      Medicago truncatula BAC Clone mth2-2298
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 115071)
AUTHORS    Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE      Direct Submission
JOURNAL    Submitted (18-OCT-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE  3 (bases 1 to 115071)
AUTHORS    Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE      Direct Submission
JOURNAL    Submitted (13-AUG-2005) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT    ----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 9337: contig of 9337 bp in length
* 9338 9437: gap of unknown length
* 9438 99732: contig of 90295 bp in length
* 99733 99832: gap of unknown length
* 99833 11207: contig of 11375 bp in length
* 11208 111307: gap of unknown length
* 111308 115071: contig of 3764 bp in length.
*
* Location/Qualifiers
* 1. 115071
* /organism="Medicago truncatula"
* /mol_type="genomic DNA"
* /db_xref="taxon:3880"
* /clone="mth2-2298"
* /clone_1kb="Medicago truncatula BAC library H2"
* 9338..9437
* /estimated_length=unknown
* 99733..99832
* /estimated_length=unknown
* 11208..111307
* /estimated_length=unknown

```

Alignment Scores: 2.2e-82 Length: 115071  
 Pred. No.: 1076.50 Matches: 196  
 Score: 74.43% Conservative: 66  
 Best Local Similarity: 55.68% Mismatches: 77

RESULT 14

Query Match: 56.72% Indels: 13  
DB: 14 Gaps: 4  
US-10-019-931-3 (1-359) x AC135566 (1-115071)

Qy 9 ILeProMetAlaIleProSerPheSerMetCysHisLeuGluLeuLeuLeuGluGly 28  
Db 31498 ATTTCATTGCTTTCACAAATTTGCCAAATCCAC----- 31533

Qy 29 LysThrArgAspValProLysAlaGluLysAspGluGly---LeuSerCysGluPheGln 47  
Db 31534 ---ACATGATATATGATGATTCATTGAGACAAAGGTCAATTCAGCAAGAAACAG 31590

Qy 48 GluMetLeuAspSerLeuProLysGluArgGlyTTPatGThraGlyTLeuTyrLeuPhe 67  
Db 31591 AATCTAATCTCTTCTCTCCAGAGAGAAAGGTGGAACACCTTAATATTTATCTATTC 31650

Qy 68 GlnGlyPheTTPCysGlnAlaLysGluIleGlnAlaIleMetSerPheGlnLysHisPhe 87  
Db 31651 CAAGGATTTGGTGCACACCGCTGAAATCCAGCATTAACACTTTCACAAAGCATTTTC 31710

Qy 88 GlnSerLeuGluAsnAspValAlaLeuAlaThrIleProLysSerGlyThrThrPhe 107  
Db 31711 CAAGCTAAGAAAGATGATGTTTGGTGCACCGTACCAAAATCAGCACAACCTGGTTA 31770

Qy 108 LysAlaLeuThrPheThrIleLeuAsnArgHisArgPheAspProValAlaSerSerThr 127  
Db 31771 AATCTCTTACCTTATGCAATATGATGATGCCAA-----AACCTTTCATTTCCCAA 31824

Qy 128 AsnHisProLeuPheThrSerAsnProHisAspLeuValProPhePheGluTyrLysLeu 147  
Db 31825 AACCATCTTGTCTTAGTTGTTCAATCCACATGATCTGTTCTTCATTGATGATACGTT 31884

Qy 148 TyrAlaAsnGlyAsp---ValProAspLeuSerGlyLeuAlaSerProArgThrPheAla 166  
Db 31885 TATGTAAGCATGACACAAATTCGATTTGCTAAATTCATGAGCCATGACCTTTGGT 31944

Qy 167 ThrHisLeuProPheGlySerLeuLysGluThrIleGluLysProGlyValLysValVal 186  
Db 31945 ACACATATCTCTTGTGACTGATGCTCAATGCAATCAAAAGTTCCAAATTCGAATATGTT 32004

Qy 187 TyrLeuCysArgAsnProPheAspThrPheIleSerSerTTPHisTyrThrAsnAsnIle 206  
Db 32005 TATATTTGTAAGAACCTTTGATGATCTTCACTTCTTTCGACCTTTGCCAACAAATTT 32064

Qy 207 LysSerGluSerValSerProValLeuLeuAspGlnAlaPheAspLeuTyrCysArgGly 226  
Db 32065 AAGCACTCTTTGGCCATCAATGAAATTTAGAGAAAGCTTTGAAATGATATGTAATGCT 32124

Qy 227 ValIleGlyPheGlyProPheTTPGluHisMetLeuGlyTyrTTPArgGluSerLeuLys 246  
Db 32125 TTAATTTGGGTTGGCCATTTTGGAAATCAATGTTAGGTTATTTGAAAGAGGCAAAAG 32184

Qy 247 ArgProGluLysValPhePheLeuArgTyrGluAspLeuLysAspAspGluIleGluThrAsn 266  
Db 32185 AGACCAAAAAATGTTCTTTTGAAGTATGAGCATGAAAGAAATTTAAATTTTAC 32244

Qy 267 LeuLysArgLeuAlaThrPheLeuGluLeuProPheThrGluGluGluLysArgLysGly 286  
Db 32245 TTGAAAAAAATGGCGTGAATTTTGGATGCTCTTTTACATTTGAGAGAGAAAGTACGGGT 32304

Qy 287 ValValLysAlaIleAlaGluLeuCysSerPheGluAsnLeuLysLysLeuGluValAsn 306  
Db 32305 GTGATTTGAAATATATTAATTAAGTTGTGTGTTTGAAGAAATGAAGGGGTGAGATTAAT 32364

Qy 307 LysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeuPheArgLysGluVal 326  
Db 32365 AATACAGAAACATTTGAAAGGACTTTGAGATTAAGTTCTTTTGAAGAAAGTAAAGTT 32424

Qy 327 SerAspTTPValAsnTyrLeuSerProSerGlnValGluArgLeuSerAlaLeuValAsp 346  
Db 32425 GGTGATTTGAGTAATATCTTTCACTTCAATGTTGAAAAAATTTCTCCAGAGTATTTGAA 32484

Qy 347 AspLysLeuGlyGlySerGlyLeuThrPheArgLeu 358  
Db 32485 GAGAAATTAGGTGATCTGATCTCAATTTTAGACTG 32520

RESULT 15  
AP008111  
LOCUS  
DEFINITION  
LOCUS corniculatus var. japonicus clone LJ105015, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 20 unordered pieces.  
AP008111  
AP008111.1 GI:56806418  
HTG, HTGS PHASE1.  
LOCUS corniculatus var. japonicus (Lotus japonicus)  
Lotus corniculatus var. japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteeae; Lotus.  
1  
Kaneko, T., Asanizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
Structural Analysis of a Lotus japonicus Genome. XI. Sequence Features and Mapping of Nine hundred twenty-one TAC Clones  
Unpublished  
2 (bases 1 to 100108)  
Sato, S.  
Direct Submission  
Submitted (26-OCT-2004) Shusei Sato, Kazusa DNA Research Institute, Department of Plant Gene Research, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan (E-mail: ssato@kazusa.or.jp, URL: http://www.kazusa.or.jp/, Tel: 81-438-52-3335 (ex. 2337), Fax: 81-438-52-3934)  
\* NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.  
1  
5202: contig of 5202 bp in length  
5203  
5302: gap of unknown length  
5303  
5566: contig of 1264 bp in length  
5567  
6666: gap of unknown length  
6667  
7760: contig of 1094 bp in length  
7761  
7860: gap of unknown length  
7861  
8624: contig of 764 bp in length  
8625  
8724: gap of unknown length  
8725  
9607: contig of 883 bp in length  
9608  
9707: gap of unknown length  
9708  
11541: contig of 1834 bp in length  
11542  
11641: gap of unknown length  
11642  
13795: contig of 2154 bp in length  
13796  
13895: gap of unknown length  
13896  
15742: contig of 1847 bp in length  
15743  
15842: gap of unknown length  
15843  
18031: contig of 3189 bp in length  
18032  
19032: gap of unknown length  
19033  
19131: gap of unknown length  
19132  
21810: contig of 2679 bp in length  
21811  
21910: gap of unknown length  
21911  
24667: contig of 2757 bp in length  
24668  
24767: gap of unknown length  
24768  
30306: contig of 5539 bp in length  
30307  
30406: gap of unknown length  
30407  
35885: contig of 6179 bp in length  
35886  
36685: gap of unknown length  
36686  
42389: contig of 5704 bp in length  
42390  
42489: gap of unknown length  
42490  
50891: contig of 8402 bp in length  
50892  
50991: gap of unknown length  
50992  
60141: contig of 9150 bp in length  
60142  
60241: gap of unknown length  
60242  
69078: contig of 8837 bp in length  
69079  
69178: gap of unknown length



*	69179	79486:	cont'g of 10310 bp in length
*	79489	79588:	gap of unknown length
*	79589	98477:	cont'g of 18889 bp in length
*	98478	98577:	gap of unknown length
*	98578	100106:	cont'g of 1531 bp in length.

Location/Qualifiers  
1. .100108

31-3 (1-359) x AP008111 (1-100108)

9 IleProMetAlaIleProSerPheSerMetCysHisIleLysLeuGluLeuLeuGly 28

63 ATTCCAATGGCACCAACAGCTCCAACCATGTTGCAAAATTCAT----- 60907

29 LysThrArgAspValProLysAlaGluGluAspGluGlyLeuSerCysGluPheGlnGlu 48

08 --TCAATGCTGATGAGGAGCAATTGATCAAGAAGAACTTAGTGAAGAATGAAGGAT 60964

49 MetLeuAspSerLeuProLysGluArgGlyTyrPargThrArgTyrLeuTyrLeuPheGln 68

610965 CTTCATTCCTCTCTCTCCAGA-GAGAGAGGTTGGAGAACCAATATCTATCTATTCCAA 6102

69 GLVPHETPCVGINAJALVSGIUILEGNAJAJEMETSERPHEGJNLVSHISPHGJN 88

61024 GGCCTTTGGTGCCAAACCAGCTGAAATCCAAGCCATAAGCACTTTTCAGAAGCACTTCCA 61083

89 SerLeuGluAsnAspValValLeuAlaThrIleProLysSerGlyThrThrTrpLeuLys 108

Db 61084 GCCAAGCACAGTGTCTTTGTTGCCACAGTGCCAAATCAGGCACCACATGTTGAAG 61143

109 AlaLeuThrphethrIleLeuAsnArchisArqPheaspProvalAlaserSerThrAsn 128

Db 61144 GCACTAACCTTTGCCATTGTAACCGCCAGAAACAT-----TTAACCTTCTCCAAGAGC 61197

129 HisProLeuPheThrSerAsnProHisAspLeuValProPhePheGluTyrIleValLeuTyr 148

Db 61198 CACCCCTTTATTAATTC CAATCCC CATGACCTTGTTCTTCATTCAGTACACTGTTAT 6125

QY 149 A1aASNGIYASP--Val1PROASPLEuSerG1YLeuAlaSerPROARgThrpheAlaThr 167

Db 61258 GGGAAACATGACAAAGGTTCTCTGATTATCCAAATCCCTGAGCCAAGGCTTTTGGACC 61317

QY 168 HisLeuProPheGlySerLeuLysGluThrIleGluLys---ProGlyValLysValVal 186

Db 61318 CACATTCATTTCTCTGCCCAATTCATCACCAGAAGCTCCAATTGCAAGATTGTT 61377

187 TYRLEUCYSARGASNPropheAspThrPheIleSerSerTPrHisTYrThrAsnAsnIle 206

Db 61378 TATATCTGCAGGAACCTTTTGACACTTTCATCTCTCTTGGTGTTTGTCAACAATA TC 61437

QY 207 LVSSerGIuSerValLeuLeuAspGlnAlaPheAspLeuTYrCysArgGly 226

Db 61438 AACCACTTCATTACCCACATTGATGCTTGAGGAGGCTTTTGACATGTTTGTAATGG 61497

QY 227 ValIleGlyPheGlyProPheTrpGluHisMetLeuGlyTyrTrpArgGluSerLeuLys 246

Db 61498 GTGGTTGGGTTTGCCCATTTTGGATCACATGTTGGTTATTGGATGAGACATTGAG 6155'

247 ArgProGluLysValPhePheLeuArgTyrGluAspLeuLysAspAspIleGluThrAsn 266

Db 61558 AGACCAACAGGTTCTTTCTTGAAATATGAGGACATGAAGAAGATCCTGTTTTCAC 6161

267 LeuLysArgLeuAlaThrPheLeuGluLeuProPheThrGluGluGluArgGly 286

Db 61618 TTGAAAAGATTGCTGAGTCTTAGGTTGCTCTTCTCTCCTTGAGGAGGAAGTGGGGT 61677

287 ValValLysAlaIleAlaGluLeuCyserPheGluAsnLeuLysLeuGluValAsn 306

Db 61678 GTATTGAGACATAATTGATTCTGTAGCTTTCAGAGATGAAGAA TTGGAGCTAAAT 6173

QY 307 LysSerAsnLysSerIleLysAsnPhcGluAsnArgPheLeuPheArgLysGlyGluVal 326

Db 61738 AGTCTGGAACATTTGCCAGGAATTCACAGACAAGTCTTGTTCAGGAAGCTGAATT 61791

327 SerAspT<sup>rp</sup>ValAsnTyrLeuSerProSerGlnValGluArgLeuSerAlaLeuValAsp 346

Db 61798 GGGATTTTGTGATCATCTGTCACTTCATGGTGGAGAAATTATCCAAAGTCATGGA 6185\*

QY 347 AspLysLeuGlyGlySerGlyLeuThrPheArgLeuSer 359

Db 61858 GAGAAGCTAGGTGGGTGTGGTTGTCAATTAGAGTGAAC 61896

Search completed: December 23, 2005, 20:41:49  
Job time : 4995 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: December 23, 2005, 16:54:45 ; Search time 604 Seconds

(without alignments)  
3961.299 Million cell updates/sec

Title: US-10-019-931-3

Perfect score: 1998

Sequence: 1 MATSMKMSIPAIRPSFSGH.....RLSALVDDKLGSGSLTRRLS 359

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 4996997 seqs, 3332346308 residues  
Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+.p2n.model -DEV=xlp  
-Q=/cgn2.1/USPRO.spool/p/US1001931/runat.16122005.165218.18429/asp.query.fasta\_1.519  
-DB=N Geneseg -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -STRAT=1 -END=1 -MATRIX=biocsum62 -TRANS=human40.cdi -LIST=45  
-DOCLALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US1001931 @CGN 1 1 1096 @runat.16122005.165218.18429 -ICPU=6 -ICPU=3  
-NO MMAP -LARGESUBSTRY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
N\_Geneseg.21.\*  
1: geneseqn1980s.\*  
2: geneseqn1990s.\*  
3: geneseqn2000s.\*  
4: geneseqn2001as.\*  
5: geneseqn2001bs.\*  
6: geneseqn2002as.\*  
7: geneseqn2002bs.\*  
8: geneseqn2003as.\*  
9: geneseqn2003bs.\*  
10: geneseqn2003cs.\*  
11: geneseqn2003ds.\*  
12: geneseqn2004as.\*  
13: geneseqn2004bs.\*  
14: geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1898	100.0	1077	4	AAF29177
2	1898	100.0	1270	3	AAC48839
3	1898	100.0	1273	3	AAC47848
4	1898	100.0	1273	12	ADQ28853

5	1898	100.0	1273	13	ADR21475	Adr21475 Arabidops
6	1553	81.8	1041	4	AAF29178	Aaf29178 Hydroxyja
7	1519	80.0	1023	12	ADQ28854	Adq28854 A. thalia
8	1519	80.0	1023	13	ADR21476	Adr21476 Arabidops
9	776	40.9	1154	3	AAC37672	Aac37672 Arabidops
10	770.5	40.6	996	6	ABZ13222	Abz13222 Arabidops
11	763.5	40.2	1069	13	ADR59777	Adr59777 Cotton cd
12	757	39.9	980	8	ADA68055	Ada68055 Arabidops
13	757	39.9	981	6	ABZ14244	Abz14244 Arabidops
14	757	39.9	981	6	ADG87892	Adg87892 A. thalia
15	757	39.9	981	6	ADG87829	Adg87829 A. thalia
16	757	39.9	981	6	ABZ42036	Abz42036 Arabidops
17	757	39.9	1043	3	AAC42382	Aac42382 Arabidops
18	738	38.9	1205	3	AAC37521	Aac37521 Arabidops
19	738	38.9	1212	3	AAC39344	Aac39344 Arabidops
20	680.5	35.9	999	8	ADA69659	Ada69659 Rice gene
21	672	35.4	1440	13	ADR60968	Adr60968 Cotton cd
22	671.5	35.4	271990	10	ADD25213	Add25213 Fertilily
23	671.5	35.4	271990	12	ADN61228	Adn61228 Radish nu
24	666.5	35.1	1251	13	ADR60969	Adr60969 Cotton cd
25	666	35.1	1358	13	ADT16337	Adt16337 plant CDN
26	664.5	35.0	1347	3	AAC37535	Aac37535 Arabidops
27	661.5	34.9	1400	8	AAD54410	Aad54410 Lolium pe
28	654	34.5	1017	10	ABZ68951	Abz68951 Nucleotid
29	653.5	34.4	1053	10	ABZ68952	Abz68952 Nucleotid
30	653.5	34.4	1053	12	ADN74550	Adn74550 Thale cre
31	653.5	34.4	1053	12	ADN74506	Adn74506 Thale cre
32	653.5	34.4	1281	3	AAC37716	Aac37716 Arabidops
33	653	34.4	1041	10	ABZ68953	Abz68953 Nucleotid
34	614	32.3	1050	12	ACL26221	Ac126221 Rice abio
35	614	32.3	1050	12	AD145084	Ad145084 Rice isop
36	613.5	32.3	2067	10	ADD25206	Add25206 Fertilily
37	613.5	32.3	2067	12	ADN61221	Adn61221 Radish nu
38	601.5	31.7	1103	14	ABZ6793	Abz6793 Rice geno
39	598	31.5	1017	8	ADA69350	Ada69350 Rice gene
40	598	31.5	1017	11	ACL26292	Ac126292 Rice abio
41	598	31.5	1017	12	AD145072	Ad145072 Rice isop
42	593.5	31.3	990	11	ACL26109	Ac126109 Rice abio
43	592	31.2	1403	13	ADX46738	Adx46738 plant ful
44	592	31.2	1526	13	ADX53055	Adx53055 plant ful
45	591.5	31.2	1044	11	ACL30321	Ac130321 Rice abio

## ALIGNMENTS

RESULT 1	
AAF29177	AAF29177 standard; DNA; 1077 BP.
XX	
XX	AAF29177;
XX	
AC	
DT	09-APR-2001 (first entry)
XX	
XX	
DE	Hydroxyjasmonic acid sulfotransferase AtST2a gene.
XX	
XX	
KW	Hydroxyjasmonic acid sulfotransferase; AtST2a; flowering time; cabbage;
KW	jaemonate; genetically modified plant; lettuce; sugar cane; carrot;
KW	increase vegetative growth; biomass increase; ds.
XX	
OS	Arabidopsis thaliana.
XX	
PN	WO200102589-A2.
XX	
PD	11-JAN-2001.
XX	
XX	
PF	06-JUL-2000; 2000WO-CA000801.
XX	
PR	06-JUL-1999; 99CA-02274873.
XX	
XX	
PA	(VARI/) VARIN L.
XX	(GIDD/) GIDDA S.
PI	Varin L, Gidda S;

XX WPI: 2001-159272/16.  
DR P-PSDB: AAB49722.

PT Methods for modulating flowering in plants, particularly useful for  
plants used in the food-processing industry, involves modifying the  
endogenous level of compounds of the jasmonate family.

PS Claim 38; Fig 7; 50pp: English.

CC This invention relates to a method for modulating flowering in a plant.  
CC The method comprises modifying the endogenous level of at least one  
CC compound of the jasmonate family in a plant. The methods are used to  
CC produce plants which are genetically modified to flower early or tardily  
CC when compared to a corresponding plant that is not genetically modified,  
CC where the modified plant has an increased (flower early) or lowered  
CC (flower tardily) level of jasmonic acid, or a compound of the jasmonate  
CC family. The method is useful for modulating flowering, particularly for  
CC plants that are used in the food-processing industry and plants with  
CC horticultural value. The method is particularly useful for e.g. delaying  
CC flowering time in crops like lettuce, cabbage, sugar cane or carrots,  
CC which results in increased vegetative growth and biomass. The present  
CC sequence represents the Arabidopsis thaliana At5g2a gene, which encodes  
CC an hydroxyjasmonic acid sulfotransferase protein, which can be used in  
CC the method of the invention

Sequence 1077 BP; 294 A; 257 C; 254 G; 272 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 2,18e-194 Length: 1077  
Score: 1898.00 Matches: 359  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-10-019-931-3 (1-359) x AAF29177 (1-1077)

QY 1 MetAlaThrSerSerMetLysSerIleProMetAlaIleProSerPheSerMetCysHis 20  
DB 1 ATGGCTACTCTCAAGCATGAGAGCATTTCCAAATGGCATCCCAAGTTTCTCCATGCTCAC 60  
QY 21 LysLeuGluLeuLeuLysGluGlyLysThrArgAspValProLysAlaGluLysAspGlu 40  
DB 61 AAGCTCGAGCTCTTAAAGAGGCAAACTCGGAGATCCCGAAAGCCGAAAGATGAA 120  
QY 41 GlyLeuSerCysGluPheGlnGluMetLeuAspSerLeuProLysGluArgGlyTTPaG 60  
DB 121 GGGCTTAAGCTGCGAGTTCCAAGAGATGATTCCTTCTTAAGAGAGAGATGAGAA 180  
QY 61 ThrArgTyrLeuTyrIleuPheGlnGlyPheTTPCyAGLAlaLysGluIleGlnAlaIle 80  
DB 181 ACTCGTACCTTACCTATTCAGAGGTTTGGTGGCCAAAGCCAAAGATTCAAGCCATC 240  
QY 81 MetSerPheGlnLysHisPheGlnSerLeuGluAsnAspValValLeuAlaThrIlePro 100  
DB 241 ATGTCTTCCAAAACATTTCCAAATCCCTCGAAAAGAGATGCTTCTCGCCACCAATACCT 300  
QY 101 LysSerGlyThrThrTyrLeuLysAlaLeuThrPheThrIleLeuAsnArgHisArgPhe 120  
DB 301 AAATCCGGTACCAACCTGGCTTAAAGCTTAACTTCAACCATCTTAAACCGTACCGGTTT 360  
QY 121 AspProValAlaSerSerThrAsnHisPheLeuPheThrSerAsnProHisAspLeuVal 140  
DB 361 GATTCGGTGGCTCGAGTACCAACCACTCTTTTACCTTCCAACTCATACCTTGTAA 420  
QY 141 ProPhePheGlnTyrTyrLeuLysTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAla 160  
DB 421 CCTTCTTGAAGTACAAAGCTTAAAGCCCAAGAGATGTTCCGATCTCGGGTCTAGCC 480  
QY 161 SerProArgThrPheAlaThrHisIleuPheProPheGlySerLeuLysGluThrIleGluLys 180  
DB 481 AGTCCAAAGAAAGCTTGGCAACCACTTACCGTTCGGTTCCCTAAAGAAAGATCGAGAA 540

QY 181 ProGlyValLysValValTyrIleuCyseArgAsnProPheAspThrPheIleSerSerTTP 200  
DB 541 CCCGGTGAAGGTCGCTACTTGTGCGGAAACCGTTTGACATTCATCTTCGCGG 600  
QY 201 HisTyrThrAsnAlaIleLysSerGluSerValSerProValLeuLeuAspGlnAlaPhe 220  
DB 601 CATTACACCAACCAATCCAGATCCAGTCAAGTCAAGCTTGTGACCAAGCTTTT 660  
QY 221 AspLeuTyrCyseArgGlyValIleGlyPheGlyProPheTTPGluHisMetLeuGlyTyr 240  
DB 661 GATCTGATTTCCGGGAGTATCGGTTTGCCCGTTTGGAACACATGTGGGATAC 720  
QY 241 TTPArgGluSerLeuLysArgProGluLysValPhePheLeuArgTyrGluAspLeuLys 260  
DB 721 TCGAGAGAGACTTGAAGACACAGAAAGTCTTTTAAAGTACGAGATCTCAA 780  
QY 261 AspAspIleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuProPheThrGlu 280  
DB 781 GACGACATCGAGACCAACTGAAAGAGCTTCACTTCTTACAGCTTCTTCCCGAA 840  
QY 281 GluGluGluArgLysGlyValValLysAlaIleAlaGluLeuCyseSerPheGluAsnLeu 300  
DB 841 GAAGGAAACGAAAGGAGTGTGAAGGCTATCCGCGAGCTGTGATGCTGAGATCTG 900  
QY 301 LysLysLeuGluValAlaLysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeu 320  
DB 901 AAGAAGTTGAGAGTGAACATGCAACAAAGTCATCAAGAACTTGAAGATCTTG 960  
QY 321 PheArgLysGlyGluValSerAspTTPValAsnTyrIleuSerProSerGlnValGluArg 340  
DB 961 TTTCCGAAAGCAAGAAAGTGAAGTGGCTTAACTTGTACCTTCACAAGTGGAAAGA 1020  
QY 341 LeuSerAlaLeuValAspAspLysLeuGluGlySerGlyLeuThrPheArgLeuSer 359  
DB 1021 TTGTCAGCTTATGATGATGACAAAGTATGATGATCTGATCTTCAAGTTGAGC 1077

RESULT 2  
AAC48839 standard; DNA; 1270 BP.  
AAC48839;  
18-OCT-2000 (first entry)  
Arabidopsis thaliana DNA fragment SEQ ID NO: 58964.  
Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway; metabolic pathway;  
KW promoter; termination sequence; ss.  
Arabidopsis thaliana.  
EP103405-A2.  
FN  
XX  
PD 06-SEP-2000.  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999;	99US-0131449P.
PR 30-APR-1999;	99US-0132048P.
PR 30-APR-1999;	99US-0132407P.
PR 04-MAY-1999;	99US-0132484P.
PR 05-MAY-1999;	99US-0132485P.
PR 06-MAY-1999;	99US-0132486P.
PR 06-MAY-1999;	99US-0132487P.
PR 07-MAY-1999;	99US-0132863P.
PR 11-MAY-1999;	99US-0134256P.
PR 14-MAY-1999;	99US-0134218P.
PR 14-MAY-1999;	99US-0134221P.
PR 14-MAY-1999;	99US-0134370P.
PR 14-MAY-1999;	99US-0134768P.
PR 18-MAY-1999;	99US-0134941P.
PR 19-MAY-1999;	99US-0135124P.
PR 20-MAY-1999;	99US-0135353P.
PR 21-MAY-1999;	99US-0135629P.
PR 24-MAY-1999;	99US-0136021P.
PR 25-MAY-1999;	99US-0136329P.
PR 27-MAY-1999;	99US-0136782P.
PR 28-MAY-1999;	99US-0137222P.
PR 01-JUN-1999;	99US-0137502P.
PR 03-JUN-1999;	99US-0137722P.
PR 04-JUN-1999;	99US-0138094P.
PR 07-JUN-1999;	99US-0138540P.
PR 08-JUN-1999;	99US-0138847P.
PR 10-JUN-1999;	99US-0139119P.
PR 14-JUN-1999;	99US-0139452P.
PR 16-JUN-1999;	99US-0139453P.
PR 17-JUN-1999;	99US-0139454P.
PR 18-JUN-1999;	99US-0139455P.
PR 18-JUN-1999;	99US-0139456P.
PR 18-JUN-1999;	99US-0139457P.
PR 18-JUN-1999;	99US-0139458P.
PR 18-JUN-1999;	99US-0139459P.
PR 18-JUN-1999;	99US-0139460P.
PR 18-JUN-1999;	99US-0139461P.
PR 18-JUN-1999;	99US-0139462P.
PR 18-JUN-1999;	99US-0139463P.
PR 18-JUN-1999;	99US-0139750P.
PR 18-JUN-1999;	99US-0139763P.
PR 21-JUN-1999;	99US-0139817P.
PR 22-JUN-1999;	99US-0139899P.
PR 23-JUN-1999;	99US-0140353P.
PR 23-JUN-1999;	99US-0140354P.
PR 24-JUN-1999;	99US-0140695P.
PR 28-JUN-1999;	99US-0140823P.
PR 29-JUN-1999;	99US-0140991P.
PR 30-JUN-1999;	99US-0141287P.
PR 01-JUL-1999;	99US-0141842P.
PR 01-JUL-1999;	99US-0142154P.
PR 02-JUL-1999;	99US-0142055P.
PR 06-JUL-1999;	99US-0142390P.
PR 08-JUL-1999;	99US-0142803P.
PR 09-JUL-1999;	99US-0142920P.
PR 12-JUL-1999;	99US-0142977P.
PR 13-JUL-1999;	99US-0143542P.
PR 14-JUL-1999;	99US-0143624P.
PR 15-JUL-1999;	99US-0144005P.
PR 16-JUL-1999;	99US-0144085P.
PR 16-JUL-1999;	99US-0144086P.
PR 19-JUL-1999;	99US-0144325P.
PR 19-JUL-1999;	99US-0144331P.
PR 19-JUL-1999;	99US-0144333P.
PR 19-JUL-1999;	99US-0144334P.
PR 19-JUL-1999;	99US-0144335P.
PR 20-JUL-1999;	99US-0144352P.
PR 20-JUL-1999;	99US-0144632P.
PR 20-JUL-1999;	99US-0144884P.

---

PR 21-JUL-1999;	99US-0144814P.
PR 21-JUL-1999;	99US-0145066P.
PR 21-JUL-1999;	99US-0145088P.
PR 22-JUL-1999;	99US-0145085P.
PR 22-JUL-1999;	99US-0145087P.
PR 22-JUL-1999;	99US-0145089P.
PR 22-JUL-1999;	99US-0145192P.
PR 23-JUL-1999;	99US-0145145P.
PR 23-JUL-1999;	99US-0145218P.
PR 23-JUL-1999;	99US-0145224P.
PR 26-JUL-1999;	99US-0145276P.
PR 27-JUL-1999;	99US-0145913P.
PR 27-JUL-1999;	99US-0145919P.
PR 28-JUL-1999;	99US-0145951P.
PR 02-AUG-1999;	99US-0146386P.
PR 02-AUG-1999;	99US-0146388P.
PR 03-AUG-1999;	99US-0147038P.
PR 04-AUG-1999;	99US-0147204P.
PR 04-AUG-1999;	99US-0147302P.
PR 05-AUG-1999;	99US-0147792P.
PR 05-AUG-1999;	99US-0147260P.
PR 06-AUG-1999;	99US-0147303P.
PR 06-AUG-1999;	99US-0147416P.
PR 09-AUG-1999;	99US-0147493P.
PR 09-AUG-1999;	99US-0147935P.
PR 10-AUG-1999;	99US-0148171P.
PR 11-AUG-1999;	99US-0148319P.
PR 12-AUG-1999;	99US-0148341P.
PR 13-AUG-1999;	99US-0148565P.
PR 13-AUG-1999;	99US-0148684P.
PR 16-AUG-1999;	99US-0149368P.
PR 17-AUG-1999;	99US-0149175P.
PR 18-AUG-1999;	99US-0149426P.
PR 20-AUG-1999;	99US-0149722P.
PR 20-AUG-1999;	99US-0149723P.
PR 20-AUG-1999;	99US-0149902P.
PR 23-AUG-1999;	99US-0149902P.
PR 23-AUG-1999;	99US-0149930P.
PR 25-AUG-1999;	99US-0150566P.
PR 26-AUG-1999;	99US-0150884P.
PR 27-AUG-1999;	99US-0151065P.
PR 27-AUG-1999;	99US-0151066P.
PR 27-AUG-1999;	99US-0151080P.
PR 30-AUG-1999;	99US-0151303P.
PR 31-AUG-1999;	99US-0151438P.
PR 01-SEP-1999;	99US-0151930P.
PR 07-SEP-1999;	99US-0152363P.
PR 10-SEP-1999;	99US-0153070P.
PR 13-SEP-1999;	99US-0153758P.
PR 15-SEP-1999;	99US-0154018P.
PR 16-SEP-1999;	99US-0154039P.
PR 20-SEP-1999;	99US-0154779P.
PR 22-SEP-1999;	99US-0155139P.
PR 23-SEP-1999;	99US-0155486P.
PR 24-SEP-1999;	99US-0155659P.
PR 28-SEP-1999;	99US-0156458P.
PR 29-SEP-1999;	99US-0156596P.
PR 04-OCT-1999;	99US-0157117P.
PR 05-OCT-1999;	99US-0157753P.
PR 06-OCT-1999;	99US-0157865P.
PR 07-OCT-1999;	99US-0158029P.
PR 08-OCT-1999;	99US-0158233P.
PR 12-OCT-1999;	99US-0158369P.
PR 13-OCT-1999;	99US-0159293P.
PR 13-OCT-1999;	99US-0159294P.
PR 13-OCT-1999;	99US-0159295P.
PR 14-OCT-1999;	99US-0159329P.
PR 14-OCT-1999;	99US-0159330P.
PR 14-OCT-1999;	99US-0159331P.
PR 14-OCT-1999;	99US-0159637P.
PR 14-OCT-1999;	99US-0159638P.

PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161932P.  
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.: 2,76e-194 Length: 1270  
Score: 1898.00 Matches: 359  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-10-019-931-3 (1-359) x AAC48839 (1-1270)

QY 1 MetAlaThrSerSerMetLysSerIlePromeAlaIleProSerPheSerMetCysHis 20  
DB 64 ATGGCTACCTCAGACATGAGACATTCCTCAATGGCGATCCCAAGTTCTCCATGTGTAC 123  
QY 21 LysLeuGluLeuLeuLysGluGlyThrArgAspValProLysAlaGluGluAspGlu 40  
DB 124 AAGCTCGAGCTCTTAAGAGAGCAAACTCCGACGTCGCCAAGCCGAGAAAGATGAA 183  
QY 41 GlyLeuSerCysGluPheGlnGluMetLeuAspSerLeuProLysGluAspGlyTrpArg 60  
DB 184 GGGCTAAGCTGGAGTCCAGAGATGTGGATCTCTTCTTAAGGAGAGAGATGGAGA 243  
QY 61 ThrArgTyrLeuTyrLeuPheGlnGlyPheTrpCysGlnAlaLysGluIleGlnAlaIle 80  
DB 244 ACTCGTTACCTTACTTACTATTCAGAGCTTTGTGTGCGAAGCCAAAGAGATTCAAGCACC 303  
QY 81 MetSerPheGlnLysHisPheGlnSerLeuGluAspAspValValLeuAlaThrIlePro 100  
DB 304 ATGCTTTCCAAAACATTCATCCATCCGAAAAACGACGTCGTTCTCGCCACCATACCT 363  
QY 101 LysSerGlyThrThrTyrLeuLysAlaLeuThrPheThrIleLeuAsnArgHisArgPhe 120  
DB 364 AATCCGGTACCAACCGGTAAAGCTTAACTTACCTTACCACTTAAACCGTACCGGTTT 423  
QY 121 AspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHisAspLeuVal 140  
DB 424 GATCCGGTTCCTTCAGATCCAAACACCCCTTTTCACTTCCAAACCTCATACCTTTGA 483  
QY 141 ProPhePheGluTyrIleLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAla 160  
DB 484 CTTTCTTCGACTACCAAGCTTAAACGCAACGAGATGTTCCGAGTCTCCGGGTCTAGCC 543  
QY 161 SerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGluThrIleGluLys 180  
DB 544 ACTCCAGAACCTTCGCAACCACTTACCGTTCGTTCCCTTAAAGAAACGATCGAGAA 603  
QY 181 ProGlyValLysValValTyrLeuCysArgAsnProPheAspThrPheIleSerSerTrp 200  
DB 604 CCCGGTGAAGGTGTGTACTTGTGTCGGAAACCGTTTGACACATCTACTCTTCGTGG 663  
QY 201 HisTyrThrAsnAsnLysSerGluSerValSerProValLeuAspGlnAlaPhe 220

DB 664 CATTACACCAACAATCAATCCAGTACGTAGACCCAGTCTTGCTAGACCAAGCTTTT 723  
QY 221 AspleuTyrCysArgGluValIleGlyPheGlyProPheTrpGluHisMetLeuGlyTyr 240  
DB 724 GATCTGTATTCGCCGGAGTGTGCGTTTGCCCGTTTGGGAACACATGTTGGGATAC 783  
QY 241 TrpArgGluSerLeuLysArgProGluLysValPhePheLeuArgTyrGluAspLeuLys 260  
DB 784 TGGAGAGAGACTTGAAAGACCAAGAAAGTCTTTTAAAGTACGAGGATCTCAA 843  
QY 261 AspAspIleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuProPheThrGlu 280  
DB 844 GACGACATCGAGACCACTTAAAGAGCGTTGCACTTTAGAGCTTCTTTACCCGAA 903  
QY 281 GluGluGluArgGlyGlyValValLysAlaIleAlaGluLeuCysSerPheGluAsnLeu 300  
DB 904 GAAAGGAAACAAAGGAGGAGTTGTGAAGCTTATCCCGACGTGTGTAGCTTGAGAAATTG 963  
QY 301 LysLysLeuGluValAsnLysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeu 320  
DB 964 AAGAACTTGGAGGTGAACATCAACATCAAGTCATCAAGAACTTGAGAAATCGATTCTTG 1023  
QY 321 PheArgLysGlyGluValSerAspTrpValAsnTyrLeuSerProSerGlnValGluArg 340  
DB 1024 TTTCCGAAAGAGAGAGTGAATGGATTGCTTAATTTGTCACTTCAAGTGGAAAGA 1083  
QY 341 LeuSerAlaLeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArgLeuSer 359  
DB 1084 TTGTCAAGCTTATGTGATGACAAAGTTAGTGTGATCTGTCTCATCTTCAAGTTAGC 1140  
  
RESULT 3  
AAC47848  
ID AAC47848 standard; DNA; 1273 BP.  
XX  
AC AAC47848;  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 55342.  
XX  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.

PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134766P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 06-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.

PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 26-JUL-1999; 99US-0145222P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149366P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151302P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 13-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155113P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0156559P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157533P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0158293P.  
PR 13-OCT-1999; 99US-0158294P.  
PR 13-OCT-1999; 99US-0158295P.  
PR 14-OCT-1999; 99US-0158329P.  
PR 14-OCT-1999; 99US-0158330P.  
PR 14-OCT-1999; 99US-0158331P.  
PR 14-OCT-1999; 99US-0158332P.  
PR 14-OCT-1999; 99US-0158637P.  
PR 14-OCT-1999; 99US-0158638P.  
PR 18-OCT-1999; 99US-0158984P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.

PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161922P.  
 PR 28-OCT-1999; 99US-0161932P.  
 PR 29-OCT-1999; 99US-0162142P.

## Alignment Scores:

Pred. No.: 2 776-194 Length: 1273  
 Score: 1898.00 Matches: 359  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 3 Gaps: 0

US-10-019-931-3 (1-359) x AAC47848 (1-1273)

OY 1 MetAlaThrSerSerMetLysSerIleProMetValIleProSerPheSerMetCysHis 20  
 DB 66 ATGGGACACTCAAGCATGAAAGCATTTCCATGGCATCCCAAGTTCTCCATGTGTCA 125  
 OY 21 LysLeuGluLeuLeuLysGluGlyLysThrArgAspValProLysValaGluGluAspGlu 40  
 DB 126 AAGCTCGAGCTCTTAAAGAGCAAACTCGGAGCTCCGAAACCCAGAAAGATGAA 185  
 OY 41 GlyLeuSerCysGluInpHeGluInpMetLeuAspSerIleProLysGluArgGlyTPArg 60  
 DB 186 GGGCTAAGCTCGAGATTCAGAGAGATGTGGATCTCTTCTTAGAGAGAGATGAGA 245  
 OY 61 ThrArgTyrLeuTyrLeuPheGlnGlyPheTyrCysGlnAlaLysGluIleGlnAlaIle 80  
 DB 246 ACTCGTTACTTACCTATTTCCAAAGGCTTTTGGTGGCCAAAGCCAAAGATTCAAGCCATC 305  
 OY 81 MetSerPheGluLysHisPheGlnSerLeuGluAsnAspValValLeuAlaThrIlePro 100  
 DB 306 ATGCTCTTCCAAAMAATTTCCATCTCGAAAGAGAGTCTTTCGCCACATACCT 365  
 OY 101 LysSerGlyThrThrTyrLeuLysAlaLeuThrPheThrIleLeuAsnArgHisArgPhe 120  
 DB 366 AATATCGGTACCAACTGGCTAAAGCTTTAATCTTCAATCTTAAACCTCATGCCGTTT 425  
 OY 121 AspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHisAspLeuVal 140  
 DB 426 GATCCGGTTGGCTCGAGTACCAACACCTCTTTCACTTCCAAACCTCATGACTTGTGA 485  
 OY 141 ProPhePheGluTyrTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAla 160  
 DB 486 CCTTCTTGAAGTACCAAGCTTTACGCCAACGGAGATGTTCCCATCTTCGGGTCTAGCC 545  
 OY 161 SerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGluThrIleGluLys 180  
 DB 546 AGTCCAAAGACGTTCCCAACCCCACTTACGTTGGTTCCTTAAAGAAACGATCGAGAA 605  
 OY 181 ProGlyValLysValValTyrLeuCysArgAsnProPheAspThrPheIleSerSerTyrP 200  
 DB 606 CCCGGGTGTAAGGTCTGTACTGTGCCGAACCCGTTTGACACATTCATCTTCTGTGG 665  
 OY 201 HisTyrThrAsnAsnLysSerGluSerValSerProValLeuLeuAspGlnAlaPhe 220  
 DB 666 CATTAACCAACAACATCAAAATCCAGTCAGTAGCCAGTCTGTAGCAACAGCTTTT 725  
 OY 221 AspLeuTyrCysArgGlyValIleGlyPheGlyProPheTyrGluHisMetLeuGlyTyr 240  
 DB 726 GATCTTATATGGCCGGGAGTGAATCGGGTTGGCCCTTTTGGAAACATGTGGGATAC 785  
 OY 241 TPArgGluSerLeuLysArgProGluLysValPhePheLeuArgTyrGluAspLeuLys 260

DB 786 TGGAGAGAGAGCTTGAAAGACACAGAGAAAGCTCTTTTAAAGGTACAGATCTCAA 845  
 OY AspAspIleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuProPheThrGlu 280  
 DB 261 GATCGATCGAGCAACCACTTGAAGGGCTGTGCAACTTCTTAGACCTTCTTCCAG 905  
 OY 281 GluGluGluArgLysGlyValValLysAlaIleAlaGluLeuCysSerPheGluAsnLeu 300  
 DB 906 GAAAGCAACGAAGAGAGTGTGTAAGGCTATCGCGAGCTGTGTAGCTTGCAGAAATCG 965  
 OY 301 LysLysLeuGluValAsnLysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeu 320  
 DB 966 AAGAGTTGAGAGTGAACAAGTCAAAAGTCGATCAAGAACTTTGAGAAATCGATTCTTG 1025  
 OY 321 PheArgLysGlyGluValSerAspThrPylAsnTyrLeuSerProSerGluValGluArg 340  
 DB 1026 TTTGGAAAGGAGAGTGAATGATTTGGTTAACTATTTGTCACTTCAACAGTGAAGA 1085  
 OY 341 LeuSerAlaLeuValAspAspLysLeuGlyLysGlyLeuThrPheArgLeuSer 359  
 DB 1086 TTGTCAAGCTTATGTGATGACAAAGTTAGTGAATCTGGTCTCATTCAAGTTGAGC 1142

## RESULT 4

ADQ28853 strand; cDNA; 1273 BP.

ID ADQ28853  
 AC ADQ28853;  
 DT 23-SEP-2004 (first entry)  
 DE A. thaliana sulfoltransferase family protein AtST2a cDNA SEQ ID NO:1.  
 KW tuber size; tuber; plant; hydroxy-jasmonate sulfoltransferase; enzyme;  
 KW jasmonate; Arabidopsis thaliana; sulfoltransferase family protein; AtST2a;  
 KW chromosome 5; gene; ss.  
 OS Arabidopsis thaliana.  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 66..1145  
 FT /tag= a  
 FT /product= "sulfoltransferase family protein AtST2a"  
 XX  
 PN MO2004057000-A2.  
 XX  
 PD 08-JUL-2004.  
 XX  
 PF 23-DEC-2003; 2003WO-CA002009.  
 XX  
 PR 23-DEC-2002; 2002CA-02415062.  
 XX  
 PA (FLOR-) FLORISYS INC.  
 XX  
 PI Varin I;  
 XX  
 DR MPI; 2004-517683/49.  
 DR P-PSDB; ADQ28855.  
 DR GENBANK; NM\_120783.  
 XX  
 PT Modulating tuber size in tuber producing plant such as Solanum tuberosum,  
 PT involves modifying in plant endogenous level of tuber modulating compound  
 PT such as jasmonate by modifying activity of hydroxy-jasmonate  
 PT sulfoltransferase.  
 XX  
 PS Claim 4; SEQ ID NO 1; 32bp; English.  
 XX  
 CC The present invention describes a method (M1) for modulating the tuber  
 CC size in a tuber producing plant. (M1) involves modifying in the plant the  
 CC endogenous level of at least a tuber modulating compound by modifying the  
 CC activity of a hydroxy-jasmonate sulfoltransferase, the tuber modulating  
 CC compound being a jasmonate. Also described: (1) a plant cell  
 CC transformation vector (1) capable of facilitating transfer and expression





PT hydroxyjasmonate by increasing in plant, level of in-vivo sulfonation of  
PT hydroxyjasmonates or decreasing level of synthesis of 11-and/or 12-  
PT hydroxyjasmonate.

XX Claim 6; SEQ ID NO 1; 299p; English.

XX The invention relates to a method for producing (M1) a male sterile  
CC plant, involving decreasing the level of 11-and/or 12-hydroxyjasmonate by  
CC increasing in the plant the level of in-vivo sulfonation of 11-and/or 12-  
CC hydroxyjasmonates or decreasing the level of synthesis of 11-and/or 12-  
CC hydroxyjasmonate. Further disclosed is a plant cell transformation vector  
CC (1) capable of facilitating transfer and expression of an exogenous  
CC nucleic acid into an isolated cell and/or facilitating integration of an  
CC exogenous nucleic acid into genome of the cell. In (M1), the level of in-  
CC vivo sulfonation of hydroxyjasmonates, is increased by increasing in the  
CC plant the endogenous activity of a hydroxyjasmonate sulfotransferase. The  
CC increasing of the level of in-vivo sulfonation of hydroxyjasmonates or  
CC the decreasing of the level of synthesis of 11-and/or 12-  
CC hydroxyjasmonate is achieved by a process chosen from genetic  
CC modification of the plant, chemical mutagenesis of the plant and  
CC selection of natural mutants, preferably genetic modification. The  
CC endogenous activity of the sulfotransferase is increased by stimulating  
CC the expression of one or more gene chosen from Arabidopsis thaliana  
CC sulfolransferase gene (AtST2a) and AtST2b. The method of the invention is  
CC useful for producing a male sterile plant. It is useful for restoring  
CC normal anther development in a genetically modified male sterile plant.  
CC The current sequence represents the Arabidopsis sulfolransferase (AtST2a)  
CC encoding gene.

XX SQ Sequence 1273 BP; 369 A; 296 C; 280 G; 328 T; 0 U; 0 Other;

XX Alignment Scores:

XX Pred. No.: 2 77e-194 Length: 1273  
XX Score: 1898.00 Matches: 359  
XX Percent Similarity: 100.00% Conservative: 0  
XX Best Local Similarity: 100.00% Mismatches: 0  
XX Query Match: 100.00% Indels: 0  
XX DB: 13 Gaps: 0

XX US-10-019-931-3 (1-359) x ADR21475 (1-1273)

QY 1 MetAlaThrSerSerMetLysSerIleProMetAlaIleProSerPheSerMetCysHis 20  
DB 66 ATGGCTACCTCAAGCATGAAGACATTCACATGCGATCCCAAGTTCTCCATGTGTCAAC 125  
QY 21 LysLeuGluLeuLeuLysGluGlyLysThrArgAspValProLysAlaGluLysAspGlu 40  
DB 126 AAGCTCGAGCTCTTAAAGAGGCAAACTCGGAGCTCCGAAAGCCGAAAGAGATGAA 185  
QY 41 GlyLeuSerCysGluIupheGluMetLeuAspSerLeuProLysGluLysArgGlyTPArg 60  
DB 186 GGGCTAAGCTCGAGTTCAGAGGAGTTCGATCTCTTCTTAAGAGAGAGATGAGAA 245  
QY 61 ThrArgTyrLeuTyrLeuPheGluGlyPheTyrCysGluAlaLysGluIleGlnAlaIle 80  
DB 246 ACTCGTTACCTTACCTATTCCAAGGGTTTGGTGCAGAGCAAGAGATTCAAGCCATC 305  
QY 81 MetSerPheGluLysIlePheGluIleSerLeuGluAsnAspValValLeuAlaThrIlePro 100  
DB 306 ATGTCTTTCCAAAACATTCATCCATCCCTCGAAAGAGAGCTCTTCGCAACATACCT 365  
QY 101 LysSerGlyThrThrTyrLeuLysAlaLeuThrPheThrIleLeuAsnArgHisATGPa 120  
DB 366 AATTCGCGTACCAACCTCGCTTAAGCTTTAATCTTCAACATCTTAAACCTCACCGGTT 425  
QY 121 AspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHisAspLeuVal 140  
DB 426 GATCCGGTGGCTCGAGTACCAACACCTCTTTCATCTTCAACCTCATGACCTTGTGA 485  
QY 141 ProPhePheGluLysTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAla 160  
DB 486 CCTTCTTGAAGTAAAGCTTTAAGCCCAAGAGAGATGTTCCATCTCTCGGGTCTAGCC 545

QY 161 SerProArgThrPheAlaThrHisLeuProPheGluSerLeuLysGluThrIleGluLys 180  
DB 546 AGTCAAGAGAGCTTCGCAACCACTTACCGTTCCGTTCCCTTAAGAGAAACGATCGAGAA 605  
QY 181 ProGlyValLysValValTyrLeuCysArgAsnProPheAsnThrPheIleSerSerTrp 200  
DB 606 CCGGCTGTGAAGGTCGTGTACTTGTGCGGAAACCGGTGTGACATTCATCTTCGTGG 665  
QY 201 HisTyrThrAsnAsnIleLysSerGluSerValSerProValLeuLeuAspGluAlaPhe 220  
DB 666 CATTAACCAACACATCAATCAATCCAGTCAAGTACCCAGCTCTTCTGACCAACCTTTT 725  
QY 221 AspLeuTyrCysArgGlyValIleGlyPheGlyProPheTyrGluIleMetLeuGlyTyr 240  
DB 726 GATCGTATTCGCGGAGTGTATCGGTTGGCCGTTTGGGAACATGTTGGGATAC 785  
QY 241 TrpArgGluSerLeuLysArgProGluLysValPhePheLeuArgTyrGluAspLeuLys 260  
DB 786 TCGAGAGAGAGCTTGAAAGACACAGAGAAAGCTTTTAAAGGTACGAGATCTCAA 845  
QY 261 AspAspIleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLysProPheThrGlu 280  
DB 846 GACGACATCGAGACCACTTGAAGAGGCTTGCACATTTCTTAAGCTTCTTCAACCGAA 905  
QY 281 GluGluGluArgGlyValValLysAlaIleAlaGluLeuCysSerPheGluAsnLeu 300  
DB 906 GAAAGGAAACGAAAGAGGTTGTGAAGCTATCGCGAGCTGTGTAGCTTCGAGAACTG 965  
QY 301 LysLysLeuGluValAsnLysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeu 320  
DB 966 AAGAGTTGGAGGTGAACAGTCAAAAGTCAATGATCAAGAAATCGATTCTTG 1025  
QY 321 PheArgLysGlyGluValSerAspTrpValAsnTyrLeuSerProSerGluLysArg 340  
DB 1026 TTTCCAAAGAGAGAGTGAATGATTCGTTAACTATTTGTCACTTCAACATGAGAAAG 1085  
QY 341 LeuSerAlaLeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArgLeuSer 359  
DB 1086 TTGTGAGCTTATGTGATGACAAAGTTAGGTGATCTGATCTCACTTCAGGTTGAGC 1142

RESULT 6  
AA29178  
ID AAF29178 standard; DNA; 1041 BP.  
XX  
AC AAF29178;  
XX  
DT 09-APR-2001 (first entry)  
XX  
DE Hydroxyjasmonic acid sulfotransferase AtST2b gene.  
XX  
KW Hydroxyjasmonic acid sulfotransferase; AtST2b; flowering time; cabbage;  
KW jasmonate; genetically modified plant; lettuce; sugar cane; carrot;  
KW increase vegetative growth; biomass increase; ds.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO200102589-A2.  
XX  
PD 11-JAN-2001.  
XX  
PF 06-JUL-2000; 2000WO-CA000801.  
XX  
PR 06-JUL-1999; 99CA-02274873.  
XX  
PA (VARI/) VARIN L.  
PA (GID/) GIDDA S.  
XX  
PI Varin L, Gidda S;  
XX  
DR WPI: 2001-159372/16.  
XX P-PSDB; AAB49723.  
PT Methods for modulating flowering in plants, particularly useful for

PT plants used in the food-processing industry, involves modifying the endogenous level of compounds of the jasmonate family.

XX Claim 38; Fig 9; 50pp; English.

CC This invention relates to a method for modulating flowering in a plant.  
 CC The method comprises modifying the endogenous level of at least one compound of the jasmonate family in a plant. The methods are used to produce plants which are genetically modified to flower early or tardily when compared to a corresponding plant that is not genetically modified, where the modified plant has an increased (flower early) or lowered (flower tardily) level of jasmonic acid, or a compound of the jasmonate family. The method is useful for modulating flowering, particularly for plants that are used in the food-processing industry and plants with horticultural value. The method is particularly useful for e.g. delaying flowering time in crops like lettuce, cabbage, sugar cane or carrots, which results in increased vegetative growth and biomass. The present sequence represents the Arabidopsis thaliana AtST2b gene, which encodes an hydroxyjasmonic acid sulfotransferase protein, which can be used in the method of the invention

XX Sequence 1041 BP; 287 A; 241 C; 242 G; 271 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	3,256-157	Length:	1041
Score:	1553.00	Matches:	299
Percent Similarity:	92.31%	Conservative:	25
Best Local Similarity:	85.19%	Mismatches:	21
Query Match:	81.82%	Indels:	6
DB:	4	Gaps:	3

US-10-019-931-3 (1-359) x AAF29178 (1-1041)

```

QY 11 MetAlaIleProSerPheSerMetCyshiSlyLeuGluLeuLeuGlyIysThr 30
DB 1 ATGGGATCCCAAGATTCTCCATGTGTCAACAAGCCCGAGCTCTTAAGAGCAAAAGC 60
QY 31 ArgAspValProLyAlaGluGluAspGlyLeuSerCyGluPhgGlnGluMetLeu 50
DB 61 -----GAAAGCCAGAAAGAGAGGCTAAGCTAAGCTTCCAAAGATGTTG 108
QY 51 AspSerLeuProLyGluArgGlyTPArgThrArgTYrLeuTYrLeuPheGlnGlyPhe 70
DB 109 GACTCTCTTCCTAAGAGAGAGAGAGCAAGAAATCGTTACCTTAATCCAAAGGTTT 168
QY 71 TrpCyGlnAlaLyGluGluGlnAlaIleMetSerPheGlnLyHisPheGlnSerLeu 90
DB 169 CGGTGCCAAGCTAAGAGATTCACAGCTTTCACAAAACATTTTCAGTCCCTT 228
QY 91 GluAsnAspValValLeuAlaThrIleProLySerGlyThrThrLeuAlaLeu 110
DB 229 CCAAGCAGAGTGTCTCGCACCAATCAATCTGGCAACAACCTGTTAAAGCTTTA 288
QY 111 ThrPheThrIleLeuAsnArgHisArgPheAspProVal---AlaSerSerThrAsnHis 129
DB 289 ACTTTCACATCTTACCCGATCGGTGATTCGGTTTCTCTATCAAGTTCCGACAC 348
QY 130 ProLeuPheThrSerAsnProHisAspLeuValProPhePheGluTYrLyLeuTYrAla 149
DB 349 CCTCTTCCACATCCACACCTCCACACCTCTTCTTCGAGTACCAAGCTTACGCC 408
QY 150 AsnGlyAspValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeu 169
DB 409 AACGGAATGTTCCGATCTCTCGGCTTACGCAAGTCCAAAGAACATTCGCAACCCACGA 468
QY 170 PropheGlySerLeuLyGluThrIleGluLySProGlyValIleValIleTYrLeuCyS 189
DB 469 CGGTCGGAGCCCTTAAGGATTCGGTCGAGATCCCAAGTGAAGTTGTGACTGTGC 528
QY 190 ArgAsnProPheAspThrPheIleSerSerTPhiSlyTYrThrAsnGlnIleYsSerGlu 209
DB 529 CGGAACCCGTTGACACATTCATCTCCATGTGGCATTTACATCAACAACATCCTCCGAG 588

```

```

QY 210 SerValSerProValLeuAspGlnAlaPheAspLeuTYrCyAspGly---ValIle 228
DB 589 TCAGTGAAGCAGATTTTCTAGACAGAACCTTTGATCTATATGTCGGGATTAAGTATC 648
QY 229 GlyPheGlyProPheThrGluHisMetLeuGlyTYrTPArgGluSerLeuYsaArgPro 248
DB 649 GGATTTGGCCCGTTTGGGAACACATGTTGGGATCTAGGAGAGAGAGCTTGAAGAGGCCA 708
QY 249 GluLySValPhePheLeuArgTYrGluAspLeuLySAspAspIleGluThrAsnLeuYs 268
DB 709 GAGAAAGCTTATTTTAAAGTACGAGATCTCAAAAGAACATGAGACCAACTTGAAG 768
QY 269 ArgLeuAlaThrPheLeuGluLeuProPheThrGluGluGluAlaArgIleValIle 288
DB 769 AAGCTAGCAAGTTTGTAGACTTCTTCCACCGAAGAGAGAACAAAGGAATTGTG 828
QY 289 LysAlaIleAlaGluLeuCySerPheGluAsnLeuLySLeuGluValIleAsnLySer 308
DB 829 AAAGCTATCGCTGATCTGTAGCTTTGAGAATCTGAAGAGTTGGAGTGAAACAAGTCA 888
QY 309 AsnLySerIleLyAsnAspGluAsnArgPheLeuPheArgLyGlyGluValSerAsp 328
DB 889 AGCAAAATGATCCAGAACTATGAGAACCGGTTCTTTAGGAAAGAGAACTGATAT 948
QY 329 TrpValAsnTYrLeuSerProSerGlnValGluArgLeuSerAlaLeuValAspAspLyS 348
DB 949 TTGGTTAATCTATTTTGTCCCATCACAAGTGAAGATGTGACCTTATGTGATACAAAG 1008
QY 349 LeuGlyGlySerGlyLeuThrPheArgLeuSer 359
DB 1009 TTAGCTGATCTGTGCTCATTGATGATTGAGT 1041

```

#### RESULT 7

ADQ28854 standard; cDNA; 1023 BP.

AC ADQ28854;

DT 23-SEP-2004 (first entry)

DE A. thaliana sulfotransferase family protein AtST2b cDNA SEQ ID NO:2.

KW tuber size; tuber; plant; hydroxy--jasmonate sulfotransferase; enzyme;

KW jasmonate; Arabidopsis thaliana; sulfotransferase family protein; AtST2b;

KW chromosome 5; gene; 88.

OS Arabidopsis thaliana.

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

Key Location/Qualifiers  
 CDS 1..1023  
 /tag=a  
 /product="Arabidopsis thaliana sulfotransferase family protein AtST2b"  
 MO2004057000-A2.  
 08-JUL-2004.  
 23-DEC-2003; 2003WO-CA002009.  
 23-DEC-2002; 2002CA-02415062.  
 (FLOR-) FLORISYS INC.  
 Varin L;  
 WPI; 2004-517683/49.  
 P-PsDB; ADO28856.  
 GENBANK; NM\_120782.  
 Modulating tuber size in tuber producing plant such as Solanum tuberosum, involves modifying in plant endogenous level of tuber modulating compound such as jasmonate by modifying activity of hydroxy-jasmonate





PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-012845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 03-MAY-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132487P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135529P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139452P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140655P.  
PR 28-JUN-1999; 99US-0140831P.  
PR 29-JUN-1999; 99US-0140931P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142350P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.

PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145085P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 06-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 09-AUG-1999; 99US-0148171P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148341P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148655P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-015138P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 11-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155559P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.

```

PR 05-OCT-1999: 99US-0157573P.
PR 06-OCT-1999: 99US-0157865P.
PR 07-OCT-1999: 99US-0158029P.
PR 08-OCT-1999: 99US-0158232P.
PR 12-OCT-1999: 99US-0158363P.
PR 13-OCT-1999: 99US-0159293P.
PR 13-OCT-1999: 99US-0159294P.
PR 13-OCT-1999: 99US-0159295P.
PR 14-OCT-1999: 99US-0159329P.
PR 14-OCT-1999: 99US-0159330P.
PR 14-OCT-1999: 99US-0159331P.
PR 14-OCT-1999: 99US-0159637P.
PR 14-OCT-1999: 99US-0159638P.
PR 18-OCT-1999: 99US-0159584P.
PR 21-OCT-1999: 99US-0160741P.
PR 21-OCT-1999: 99US-0160767P.
PR 21-OCT-1999: 99US-0160768P.
PR 21-OCT-1999: 99US-0160770P.
PR 21-OCT-1999: 99US-0160814P.
PR 21-OCT-1999: 99US-0160815P.
PR 22-OCT-1999: 99US-0160980P.
PR 22-OCT-1999: 99US-0160981P.
PR 22-OCT-1999: 99US-0160989P.
PR 25-OCT-1999: 99US-0161404P.
PR 25-OCT-1999: 99US-0161405P.
PR 25-OCT-1999: 99US-0161406P.
PR 26-OCT-1999: 99US-0161359P.
PR 26-OCT-1999: 99US-0161360P.
PR 26-OCT-1999: 99US-0161361P.
PR 28-OCT-1999: 99US-0161920P.
PR 28-OCT-1999: 99US-0161992P.
PR 28-OCT-1999: 99US-0161993P.
PR 29-OCT-1999: 99US-0162142P.

```

## Alignment Scores:

```

Pred. No.: 2,22e-73 Length: 1154
Score: 776.00 Matches: 143
Percent Similarity: 61.498 Conservative: 71
Best Local Similarity: 41.098 Mismatches: 122
Query Match: 40.894 Indels: 12
DB: 3 Gaps: 4

```

US-10-019-931-3 (1-359) x AAC37672 (1-1154)

```

QY 17 SerMetCysHisIleYs-----LeuGluLeuLeuLeuGluGluYs 29
DB 57 TCTTTGTTGTTTCAAAATCGTGGAAACAATTATCAACAATGAGGCTTTCTAAAGAA----- 110
QY 30 ThrArgAspValProLysAlaGluGluAspGluGluLeuSerCysGluPheGlnGluMet 49
DB 111 GCTCATCAACCTTCCAAACTATCATGAAAGACGACAAAGTTAGTCAGAAACCAAGAACTTG 170
QY 50 LeuAspSerLeuProLysGluArgGlyTTPArgThrArgTyrLeuPheGlnGly 69
DB 171 ATCACTTCTCTACCTTCAGACAAAGATTTCATGCGGTATGCTCTACAACTACAAAGGT 230
QY 70 PheTTPCysGlnAlaLysGluLeuGlnAlaIleMetSerPheGlnLysHisPheGlnSer 89
DB 231 TGTGTGTAATCTCAAAACACACTCCAAAGCGCTTCTTGAAGTCCAAACCACTTCAAGCCA 290
QY 90 LeuGluAsnAspValValLeuAlaThrIleProLysSerGlyThrThrTrpLeuYsAla 109
DB 291 CGAGATACCTGAATATATCTCGCTCTTGTGCCAAAGGTGGAACCACTTGGCTCAAAATCC 350
QY 110 LeuThrPheThrIleLeuAsnArgHisArgPheAspProValAlaSerSerThrAsnHis 129
DB 351 CTAAATTTTCGCTGTTCTACATGAGAAAGTACCGC-----GAAACCCCTCAACACACT 404
QY 130 ProLeuPheThrSerAsnProHisAspLeuValProPhePheGluTyrLysLeuYsAla 149
DB 405 CTTTGGCTCTTCAAAAACCTTCATGACCTTGTCCCATTTCTTGAGGTATGATTATCGCT 464
QY 150 AsnGlyAspValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisIleu 169

```

```

DB 465 ANTACCCAAATTCGCGATCTCGCAAAAGTATTTCTCTATGATCTTTTCTACACAGAG 524
QY 170 PropheGlySerLeuLysGluThrIleGluLysProGlyValLysValYrLeuCys 189
DB 525 CACTTACAAAGATTCGTGGAAAGCCACCAAAA-----GCTGCAAAAACGTATATGCTGT 581
QY 190 ArgAsnProPheAspThrPheIleSerSerTrpHisTyrThrAsnAsnLysSerGlu 209
DB 582 AAGAGTATCAAGATACGTTGTCTCGCGTGGCATTTAGAAACATGTTGATCGCACCC 641
QY 210 SerValSerProValLeuLeuAspGlnAlaPheAspLeuYrCysArgGlyValIleGly 229
DB 642 AAGATGATCAAGCACTTTTGAAGCTCATGTTGATGCTTATGTTAGAGAGATTTCTCTTA 701
QY 230 PheGlyProPheTTPGluHisIleMetLeuGlyTTPArgGluSerLeuLysArgProGlu 249
DB 702 TATGAGACCTTATGGAACATGATATTTGAGCTATTGGAAGGAGCTTGGAAACCAAGAG 761
QY 250 LysValLysPheLeuArgTyrGluAspLeuYsAspAspIleGluThrAsnLeuYsArg 269
DB 762 AATGTTCTTTTCATAGATGACGAAAGATTAATTAGAGAGCCTCGTGTTCAACTCAAGAGA 821
QY 270 LeuAlaThrPheLeuGluLeuProPheThrGluGluGluGluArgLysGlyValYsAla 289
DB 822 CTCGCCGAGTTCTTGGAAATGTCATTCACCAAGAAAGAAAGAAAGTGCATCGGTGAG 881
QY 290 AlaIleAlaGluLeuCysSerPheGluAsnLeuYsLysLeuGluValAsnLysSerAsn 309
DB 882 GAGATCTTGAAGTGTGTAGTTATTAACAATTATTAACAATTGGAAGTTATTAAGAATGG 941
QY 310 LysSerIleLysAsnPheGluAsnArgPheLeuPheAlaGlyLysGlyValSerAspTTP 329
DB 942 ACAACGAGAAATGCTGATGATCTCGAGGTCTTCTAGAAAGGTAAGTGTGTGATTTGG 1001
QY 330 ValAsnTyrLeuSerProSerGlnValGluArgLeuSerAlaLeuValAspAspLysLeu 349
DB 1002 AAGAAATCTTTCATCCCAAAATGCGCAAAACCTTTGATGAGATTAATTGACTATGACTTA 1061
QY 350 GlyGlySerGlyLeuThrPheArg 357
DB 1062 GGAGACTCCGGTTTGATATTTCAA 1085

```

RESULT 10

```

AB213222
ID AB213222 standard; DNA; 996 BP.
AC AB213222;
DT 21-JAN-2003 (first entry)
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 1027.
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
OS Arabidopsis thaliana.
PN WO200216655-A2.
PD 28-FEB-2002.
PF 24-AUG-2001; 2001WO-US026685.
PR 24-AUG-2000; 2000US-0227866P.
PR 26-JAN-2001; 2001US-0264647P.
PR 22-JUN-2001; 2001US-0300111P.
PA (SCRI ) SCRIPPS RES INST.
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
PI Harper JF, Krepe J, Wang X, Zhu T;
DR MPI; 2002-304127/34.

```





comprises transforming a plant with a recombinant construct comprising a promoter region functional in a plant cell operably joined to a polynucleotide comprising a coding sequence for a polypeptide associated with the property, and growing the transformed plant. The polypeptide is useful for improving plant cold tolerance, manipulating growth rate in plant cells by modification of the cell cycle pathway, improving plant drought tolerance, providing increased resistance to plant disease, producing galactamman (or laminin or plant growth regulators), improving plant heat tolerance, improving plant tolerance to herbicides, increasing the rate of homologous recombination in plants, improving plant tolerance to extreme osmotic conditions or to pathogens or pests, improving yield by modification of photosynthesis, modifying seed oil or protein yield and/or content, improving yield by modification of carbohydrate, nitrogen or phosphorus use and/or uptake, or improving yield by providing improved plant growth and development under at least one stress condition. The polynucleotide and polypeptide may also be used in recombinant DNA constructs, in physical arrays of molecules, as plant breeding markers, or in computer-based storage and analysis systems. The present sequence is a Cotton plant cDNA of the invention. NOTE: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at [seqdata.uspto.gov/sequence.html?DocID=D-20040181930](http://seqdata.uspto.gov/sequence.html?DocID=D-20040181930). However only 6585 polynucleotide sequences were available, the remaining 52213 polynucleotides and all 58798 protein sequences were not present.

SQ Sequence 1069 BP; 318 A; 252 C; 210 G; 289 T; 0 U; 0 Other;

**Alignment Scores:**

Pred. No.:	4,43e-72	Length:	1066
Score:	763.50	Matches:	139
Percent Similarity:	64.17%	Conservative:	67
Best Local Similarity:	43.30%	Mismatches:	110
Query Match:	40.23%	Indels:	5
DB:	13	Gaps:	2

US-10-019-931-3 (1-359) x ADR59777 (1-1069)

Oy		12	AlaIleProSerPheSerMetCysHisIlySleuGlueLeuLeuIleuGluGlySerThrArg	31
Db		119	TCAITTAACAGCGTGAATAATGATGGCCAT-----ACTCAAGCTAAGATTCACTTTCT	169
Oy		32	AspValProLysalaGluGluuAspGluGlyLeuSerCysGluPhedngluMetLeuAsp	51
Db		170	GATGTTCCCAATATACCTGCTAGAATTGGGTCTCACCCAGAATGCAGACGTTATATCC	229
Oy		52	SerLeuProLysgluIargGlyTrpArgThrAArgIyrLeuTyrlleuPhedngInyPheTrp	71
Db		230	TCACGCCAGCGCAGAAAGGTTGGTGTCCAATCTCTCATCAATATACAGGGGTTTTGG	289
Oy		72	CysGlnalIylseGluileginalaIleMetSerPheGlnIlyshiIspheGlnserLeuGlu	91
Db		290	CACACCACCTCGGCAGATTACAACAGAGTGCTTACTCTGCCAAAACACCTTCAAGCTCAGAG	349
Oy		92	AsnAspValValLeuualathrIlePolyserylThrThrTrrleuLySalaleuthr	111
Db		350	ACAGATATCTCTCTTGTTAACCACTCCAAATCAGGAACAACAGTGTAAAGGCTATATGTT	409
Oy		112	PheThrIleLeuasnArgHisArgPheasPProvalaIaserSerthrAshniIspoleu	131
Db		410	TTTGCGCTTGATTAACCGATCAAGTATCCC-----AACACTGATAACATCAACCTTTG	463
Oy		132	PheThrSerAsnProHisAspleuValProphePheGluTyrlsLeuTyrylaaengly	151
Db		464	CTCTCAGAGAACCCCTCACATCTTGTTCCATTTTGGAGTATGGCTTTTACATTTGATAGT	523
Oy		152	AspValProAspleuSerGlyLeuualaserProArgThrPhelalaphrIsleuProPhe	171
Db		524	CAGGTTCCTTAACCTTCAACCACTTTATCATCTCCAGGCTATTCGGAATCATTTACCCCTT	583
Oy		172	GlySerLeuLyeglUthrIleGluYsbProGlyValIysValValTyrlleuCySarGaen	191
Db		584	GTTTCATTGCGCAAGATCGGCAAGAACTCATCTTGCACACTGTGTTATTTATATAGAAAC	643

QY	192	ProPheasprhrPheIIeSerSetrPhIaIeTyrThrAsnIleuysSerGluSerVal	211
Db	644	CCAAAGATACCTTTCGATCGCTTTGGCACTTACGAAACAAGCTGAACAACCAAGATATG	703
QY	212	SerProValIleuIeuaSPGInAlaPheAspLeuTyrCyAargGlyValIleGlyPheGly	231
Db	704	GGAAGCACTCTCTTGAAGAGACCTTGCATTAAGTTTCCACAGCGCTGAGTTGTATGGA	763
QY	232	ProPheTrpGlnIAsmetLeuGlyTyrTrpArgIuSerLeuysbArpProGluysVal	251
Db	764	CCGTTTTGGGACCATGTTTTATGTTATTTGGAACAAGACTTGGAAAAACCTGAAGAAGTC	823
QY	252	PhePheLeuArgrLyrgIuAspLeuLyAspAspIleGluThrAsnLeuLybArGleuAla	271
Db	824	TTGTGTTTGAATAATCGAGAAATGMAAGAACAGCCACACTTCAGCTGAGTGAAGCTGCT	883
QY	272	ThrPheIuGluIueuProPheThnGluGluGluIuArgrLyGlyValValIyAlaIle	291
Db	884	CAGTTCCTCGGATGCCATTTTCCATGAAAGAAAGACTGATGTGCGGTGGATGGTATA	943
QY	292	AlaIeIueCySerPheGluAsnIleuLybIyLeuGluValAlAsnLybSerAsnLySer	311
Db	944	CAAAATCTATGACGCTTTGAGANTTTGACCAACTGGACGTTAAACATATACCCGAAGTTG	1003
QY	312	IleLyAsnPheGluAsnArgrPheLeuPheArgrLyGlyIuValSerAspTrpValAsn	331
Db	1004	GCATCGGGTGAGGTTTCAAGGACATTTTTCACCAACGTGATAATTTGGAGATCTCTTAACAT	1063

## RESULT 12

ID ADA68055 standard; DNA; 980 BP.

AC ADA68055;  
XX  
XX 20-NOV-2003 (first entry)  
DT  
XX  
XX Arabidopsis thaliana gene, SEQ ID 310.  
DE  
XX  
XX Plant; bacterial infection; fungal infection; viral infection; ds.  
KM  
XX  
XX Arabidopsis thaliana.  
OS  
XX  
XX WO200300898-A1.  
PM  
XX  
XX 03-JAN-2003.  
PD  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
PF  
XX  
XX 22-JUN-2001; 2001WO-IB001105.  
PR  
XX  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
PA  
XX  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,  
PI Karagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
PI  
XX  
XX WPI; 2003-1/5290/17.  
DR  
XX  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
PT  
XX  
XX Claim 6; SEQ ID NO 310; 899p; English.  
PS  
XX  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC expression of the gene in an uninfected plant, in a mutant plant that





```

US-10-019-931-3 (1-359) x ADG87829 (1-981)
QY 33 ValProlybAlaGluGluAspGluLeuSerCysGluPheGlnGluMetLeuAspSer 52
DB 19 GTTCTGCTTACTTGGAGATGAGATCTGACACAAAGAAACAGCTCTGATCTCTCT 78
QY 53 LeuProlybGluArgGlyThrArgThrArgGlyLeuLeuPheGlnGlyPheTrpCys 72
DB 79 CTCTCTAAAGAAAGAGTTGGTTAGTGAAGTAAATATGAAATTCACAGACTTGGCAC 138
QY 73 GlnAlaIysGluIleGlnAlaIleMetSerPheGlnIysPheGlnSerLeuGluAsn 92
DB 139 ACACAGAGCATTTTACAGAGATCTTGATCTGCAAAAACGGTTTGAAGCTAAAGATTCC 198
QY 93 AspValValLeuAlaThrIleProlybSerGlyThrIleTrpLeuValAlaLeuThrPhe 112
DB 199 GACATTATCTCGTCACTAATCTTAATCAGGTACACACTGGTATTAACCTCTTGCTTT 258
QY 113 ThrIleLeuAsnArgHisArgPheAspProValAlaSerSerThrAsnHisProLeuPhe 132
DB 259 GCTCTCTTAACCGACAAAGTTT---CCAGTTCTTCTTCTGGTACCATCTCTCTTCTG 315
QY 133 ThrSerAsnProHisArgPheValProPhePheGluTrpIleLeuValAlaAsnGlyAsp 152
DB 316 GTACCAATCCACACCTTCTTGCCCTTCTGGAAGAGATTACTACGAGTCCACAGAT 375
QY 153 ValProAspLeuSerGlyLeuAlaSerProArgTrpPheAlaThrHisLeuProPheGly 172
DB 376 TTC---GATTCTCCAGTTGCTTCCAGCTGATGAAACGACATGCGCATCTT 432
QY 173 SerLeuIysGluThrIleGluIysProGlyValIleValIleValIleGlyPheGlyPro 192
DB 433 TCGCTCCCGAGTCTTAAAGACTGCTGTGTAAGATTGTGATGTTGTGGAACCTT 492
QY 193 PheAspThrPheIleSerSerTrpHisArgThrAsnAsnIleIysSerGluSerValSer 212
DB 493 AAGGACATGTTTGTGCTTATGCGCATTTGGAAAAGCTAGCTCTCGTGAACCGCG 552
QY 213 ProValLeuLeuAspGlnAlaPheAspLeuTrpCysArgGlyValIleGlyPheGlyPro 232
DB 553 GATTATCTATCGAAAACCGGTTGAAGCGTTTGTGAAGGGAAGTTTAAGGTGACCC 612
QY 233 PheTrpGluHisMetLeuGlyTrpTrpArgGluSerLeuIysArgProGluIysValPhe 252
DB 613 TTTTGGGATCATATATATGAGATCTGTATGCAAGCCGAGAAATCCAAAGAGCTTGG 672
QY 253 PheLeuArgTrpGluAspLeuIysAspAspIleGluThrAsnLeuIysArgLeuAlaThr 272
DB 673 TTTGTTACTTACGAGAGCTAAAGAGACAGACCGAAGTTGAGATGAAGCGGAG 732
QY 273 PheLeuGluLeuProPheThrGluGluGluGluArgGlyValIleValIleAlaIle 292
DB 733 TTTCTGGAATGTGGCTTTATTTGAAGAGAAAGAA-----GTGAGAGATTTGG 780
QY 293 GluLeuCysSerPheGluAsnLeuIysIleValIleAsnIysSerAsnIysSerIle 312
DB 781 AAGTTCTAGCTTAAAGCTTAACTAATTTGGAAGTTTAAACAAAGGAAATTTGCCA 840
QY 313 LysAsnPheGluAsnArgPheLeuPheArgGlyGlyValIleSerAspTrpValAsnTrp 332
DB 841 AATGGAATAGAGCTAAAACTTTCTTAGAAAAGAGAGATTTGAGAGATGAGAGATCT 900
QY 333 LeuSerProSerGlnValGluArgLeuSerAlaLeuValAspAspIysLeuGlyIysSer 352
DB 901 TTGAGTGAAGCATTTGGCAGAGAAATTTAGTAACAATTTGAAGAGATTAAAGTTCT 960
QY 353 GlyLeuThrPhe 356
DB 961 GGTCTTAATTT 972

```

RESULT 15  
 ADG87829  
 ID ADG87829 standard; cDNA; 981 BP.

```

XX AC ADG87829;
XX XX 22-APR-2004 (first entry)
XX DT
XX XX
XX DE A. thaliana RPP7/RPP8-upregulated pathogen infection-related gene #271.
XX XX pathogen infection-related gene; plant; Peronospora parasitica;
XX KM defence mechanism; RPP7; RPP8; pathogen resistance; transgenic plant;
XX KM oomycete; fungus; bacterium; virus; nematode; insect; aphid; gene; ss.
XX XX Arabidopsis thaliana.
XX OS
XX XX
XX PN MO200222675-A2.
XX PD
XX XX
XX PF 14-SEP-2001; 2001WO-US028506.
XX PR 15-SEP-2000; 2000US-0232778P.
XX PR 22-JUN-2001; 2001US-0300183P.
XX XX
XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX PA (UNNC-) UNIV NORTH CAROLINA.
XX PA (GLAZ/) GLAZEBROOK J.
XX PA (MANG/) WANG X.
XX PA (DANG/) DANG L.
XX PA (EUG/) EUGEN T.
XX PA (ZHU/) ZHU T.
XX PI Glazebrook J, Wang X, Dang L, Eugen T, Zhu T;
XX DR MPI; 2002-292409/33.
XX PT
XX PT Novel isolated polynucleotide, useful for conveying pathogen resistance
XX XX to plants, and for identifying plants infected with a pathogen.
XX PS
XX PS Claim 3; SEQ ID NO 271; 605bp; English.
XX XX
XX CC The invention relates to 691 Arabidopsis thaliana genes (ADG87559--
XX CC ADG87557) whose expression is altered in response to pathogen infection,
XX CC and to homologues of these genes from other plants or fungi, especially
XX CC from maize, soybean, barley, alfalfa, sunflower, canola (oilseed rape),
XX CC cotton, peanut, sorghum, tobacco, sugarbeet, rice or wheat. The
XX CC expression of genes of the invention was upregulated or downregulated in
XX CC Arabidopsis plants infected with the oomycete Peronospora parasitica,
XX CC indicating that they play a role in defence mechanisms. The genes of the
XX CC invention are regulated by RPP7 or RPP8 which act via unconventional
XX CC signalling cascades, or by the RPP4-dependent pathway. The invention also
XX CC relates to polypeptides encoded by the pathogen infection-related genes;
XX CC promoter motifs from pathogen infection-related genes (ADG88243-ADG88327)
XX CC; expression cassettes, host cells and pathogen-resistant transgenic
XX CC plants and their progeny, comprising a polynucleotide of the invention;
XX CC and a method of identifying a plant cell infected with a pathogen. The
XX CC polynucleotide sequences and methods of the invention are useful for
XX CC identifying plants infected with a pathogen, and for conferring
XX CC resistance to pathogens such as oomycetes, fungi, bacteria, viruses,
XX CC nematodes and insects (e.g., aphids). The present sequence represents an
XX CC Arabidopsis thaliana gene whose expression is altered in response to
XX CC Peronospora parasitica infection. Note: The sequence data for this patent
XX CC did not form part of the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX XX
XX SQ Sequence 981 BP; 283 A; 178 C; 228 G; 292 T; 0 U; 0 Other;

```

## Alignment Scores:

Pred. No.:	1,97e-71	Length:	981
Score:	757.00	Matches:	148
Percent Similarity:	61.11%	Conservative:	50
Best Local Similarity:	45.68%	Mismatches:	120
Query Match:	39.88%	Indels:	6
DB:	6	Gaps:	3

US-10-019-931-3 (1-359) x ADG87829 (1-981)

Job time : 620 secs

```
OY 33 ValProLybAlaGluGluAspGluLeuSerCyseGluPheGlnGluMetLeuAspSer 52
DB 19 GTTCCTGCTTACTTGGAGATGATGAGATCTGACACAAAGAAAGAGCTGTGATCTCTT 78
OY 53 LeuProLybGluArgGlyTTPArgThrArgTyrLeuTyrLeuPheGlnGlyPheTyrCys 72
DB 79 CTTCTTAAGAGAAAGTTGTTAGTGAAGTAATATGAAATCCAGAGACTTGGCAC 138
OY 73 GlnAlaLySGluIleGlnAlaIleMetSerPheGlnLyseHspheGlnSerLeuGluAsn 92
DB 139 ACACAAAGCTATTTCACAGAAATCTTGATCTGCCAAAACGCTTGAGACTTAAGATTCC 198
OY 93 AspValValLeuAlaThrIleProLybSerGlyTyrThrTyrLeuLybAlaLeuThrPhe 112
DB 199 GACATTATCTCTGTCATCAATCAATCAAGTACCACTGGTTAAAGCTTGTCTT 258
OY 113 ThrIleLeuAsnArgHisArgPheAspProValAlaSerSerThrAsnHisProLeuPhe 132
DB 259 GCTCTCTTAACCGACACAGATT--CCAGTTCTTCTTCTGTAAACCATCTCTCTG 315
OY 133 ThrSerAsnProHisAspLeuValProPhePheGluTyrLyseLeuTyrAlaAsnGlyAsp 152
DB 316 GTACACCAATCCACACTTCTGTGCTCTTCTGGAAGAGTTTACTACAGATCCCAAGAT 375
OY 153 ValProAspLeuSerGlyLeuAlaSerProArgTyrPheAlaThrHisLeuProPheGly 172
DB 376 TTC--GATTCTCCAGTTTGCTTCTCCAGACATGATGACACGACATATCGCATTT 432
OY 173 SerLeuLybGluThrIleGluLybProGlyValLybValValTyrLeuCybArgAsnPro 192
DB 433 TCGCTCCCGAGTCGTGAAGAGCTGCTGTGAAGATTGTATGTGTAGGAACCT 492
OY 193 PheAspThrPheIleSerSerThrHisTyrThrAsnAsnIleLybSerGluSerValSer 212
DB 493 AAGGACATGTTGTGTCTTATGCAATTTTGGAAAAAGCTAGCTCTGAGAAACCGCG 552
OY 213 ProValLeuLeuAspGlnAlaPheAspLeuTyrCybArgGlyValIleGlyPheGlyPro 232
DB 553 GATTATCTCATCGAAAAAGCGGTGAGCGTTTGTGAAGGAGATTATAGTGGAACCC 612
OY 233 PheTyrPGluHisMetLeuGlyTyrTTPArgGluSerLeuLybArgProGluLybValPhe 252
DB 613 TTTTGGGATCATATATGTGAGTACTGTATGCAAGCCGCGAGATCCGAAACAGTCTTG 672
OY 253 PheLeuArgTyrGluAspLeuLybAspAspIleGluThrAsnLeuLybArgLeuAlaThr 272
DB 673 TTTGTTACTTACGAGAGACTAAGAACGACGACGAAAGTTGAGATGAAAGCGGATCGCGAG 732
OY 273 PheLeuGluLeuProPheThrGluGluGluArgLybGlyValValIleAlaIleAla 292
DB 733 TTTCTTGAATGTGGCTTTATGAGAGAAAGAA-----GTGAGAGAGATTGTG 780
OY 293 GluLeuCybSerPheGluAsnLeuLybLybLeuGluValAlaAsnLybSerAsnLybSerIle 312
DB 781 AAGTGTGATGCTTGAAGCTTGAATGATTAATTGGAAGTTAACAAAGAGGAAATTGCCA 840
OY 313 LysAsnPheGluAsnArgPheLeuPheArgLybGlyGluValSerAspTyrValAsnTyr 332
DB 841 AATGGAATAGACATAAACTTCTTTAGAAAAGAGAGATTGAGAGATGAGAGATACT 900
OY 333 LeuSerProSerGlnValGluArgLeuSerAlaLeuValAspAspLybLeuGlyLybSer 352
DB 901 TTGAGTGAATCATTTGGCAGAGAAATTGATAGAACCATTGAAGAGAGATTAAAGTTCT 960
OY 353 GlyLeuThrPhe 356
DB 961 GGTTCTTAATTT 972
```

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: December 23, 2005, 18:49:27 ; Search time 3717 Seconds  
(without alignments)  
4518.849 Million cell updates/sec

Title: US-10-019-931-3

Perfect score: 1898

Sequence: 1 MATSSKSIIMAIPISEFMCH.....RLGALVDDKIGSGSLTRRLS 359

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=x1p  
-Q=/cgr21/USPTO.spool/p/US10019931/runat.16122005.165219.18452/app\_query.fasta.1.519  
-DB=EST -OPMT=fastap -SUPFIX=p2n.rst -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pct -NORM=ext -HEAFCSTZ=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US10019931 @CGN 1.1 8010 @runat.16122005.165219.18452 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEOBURY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -XLANGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : EST:\*

1: gb\_eest1.\*  
2: gb\_eest2.\*  
3: gb\_eest3.\*  
4: gb\_hcc.\*  
5: gb\_eest4.\*  
6: gb\_eest5.\*  
7: gb\_eest6.\*  
8: gb\_eest7.\*  
9: gb\_gss1.\*  
10: gb\_gss2.\*  
11: gb\_gss3.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	1877	98.9	4	CNSOANWE BX832510 Arabidops
2	1791.5	94.4	4	CNSOANWZ BX831873 Arabidops
3	1778	93.7	4	CNSOANOV BX831865 Arabidops
4	1582	83.4	4	CNSOAI86 BX829897 Arabidops
5	1573	82.9	4	CNSOAIK2 BX829446 Arabidops
6	1533	80.8	4	CNSO9YST BX831024 Arabidops
7	1528	80.5	4	CNSOANWP BX831860 Arabidops

c	8	1103.5	58.1	757	9	BZ062934	BZ062934 11c02b11.
	9	1094	57.6	811	9	BH733391	BH733391 BOMIUC6TR
	10	968	51.0	727	6	CD837475	CD837475 BM45.052G
	11	936.5	49.3	734	6	CD834047	CD834047 BM45.040G
	12	916.5	48.3	771	1	AU238005	AU238005 AU238005
	13	902	47.5	875	7	CV130872	CV130872 X9SP07a08
	14	887.5	46.8	817	8	CK043920	CK043920 UCRCS07.1
	15	870	45.8	574	1	AV827990	AV827990 AV827990
	16	854	45.0	523	1	AV442770	AV442770 AV442770
	17	854	45.0	635	6	CB260303	CB260303 81-E9537-
	18	853	44.9	651	6	CD833738	CD833738 BM45.001G
	19	847.5	44.7	717	8	CK046401	CK046401 UCRCS07.7
	20	830.5	43.8	782	1	CN189921	CN189921 UCRCS06.0
	21	819	43.2	657	6	CD822905	CD822905 BM25.047A
	22	804	42.4	641	1	AV520946	AV520946 AV520946
	23	798	42.0	585	6	CB260353	CB260353 77-E01117
	24	798	42.0	638	1	AV441110	AV441110 AV441110
	25	797.5	42.0	808	7	CN782243	CN782243 EST00339
	26	797	42.0	760	9	CC952179	CC952179 BOIGS48TF
	27	790.5	41.6	795	6	CA799562	CA799562 sat35e10.
	28	785.5	41.4	531	1	AV442439	AV442439 AV442439
	29	777	40.9	738	6	CF418052	CF418052 USDA-PP_1
	30	770.5	40.6	781	6	CA782758	CA782758 sat54b11.
	31	757	39.9	889	7	CV236747	CV236747 WS01225.B
	32	755	39.8	1274	4	CNSOANV7	DR927927 EST11946
	33	750.5	39.5	698	8	DR927927	DR927927 EST11946
	34	745	39.3	698	5	BH831310	BH831310 T020A09.P
	35	733.5	38.6	847	7	CO077215	CO077215 GR_Ea380
	36	729.5	38.4	849	7	CO120965	CO120965 GR_EB024
	37	727.5	38.3	934	10	CG919663	CG919663 MBET65TR
	38	725.5	38.2	701	6	CF509356	CF509356 USDA-PP_1
	39	724	38.1	694	5	BH834467	BH834467 T061B09.P
	40	719	37.9	855	10	CG950535	CG950535 MRECO3TF
	41	716	37.7	680	6	CB006702	CB006702 VYCO36H04
	42	707.5	37.3	719	8	DT043830	DT043830 Mdtrp1007
	43	681	35.9	916	8	CK547767	CK547767 gmrtdrNS0
	44	680.5	35.9	999	10	CL954941	CL954941 OsIRUN005
	45	678.5	35.7	832	8	CX192873	CX192873 38-E02278

## ALIGNMENTS

RESULT 1	CNSOANWE	1396 bp	mRNA	linear	HTC 06-FEB-2004
LOCUS	Arabidopsis thaliana Full-length cDNA Complete sequence from clone				
DEFINITION	GSJLPGH78ZAO4 of Hormone Treated Callus of strain col-0 of Arabidopsis thaliana (thale cress).				
ACCESSION	BX832510	GI:42458247			
VERSION	BX832510.1				
KEYWORDS	HTC; GSJL cDNA.				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsie.				
AUTHORS	Castelli,V., Aury,J.M., Jallou,O., Wincker,P., Clepet,C., Menard,M., Gruaud,C., Querier,F., Scarpelli,C., Schachter,V., Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.				
TITLE	Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences: A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1396)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqre@genoscope.cns.fr				
COMMENT	- Web : www.genoscope.cns.fr The sequences are based on single pass reads. Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.				



FEATURES		Location/Qualifiers	
source		1..1201	
		/organism="Arabidopsis thaliana"	
		/mol_type="mRNA"	
		/db_xref="taxon:3702"	
		/clone="GSLTPGH432B05"	
		/tissue_type="Hormone Treated Callus"	
		/ecotype="Col-0"	
		/plasmid="pCMTSPORT 6"	
		complement(1..1201)	
		/gene="At5g07010"	
		/gene="At5g07010"	
ORIGIN			
Alignment Scores:			
Pred. No.:		7.81e-195	Length: 1201
Score:		1791.50	Matches: 348
Percent Similarity:		96.41%	Conservative: 1
Best Local Similarity:		96.13%	Mismatches: 10
Query Match:		94.39%	Indels: 3
DB:		4	Gaps: 1
US-10-019-931-3 (1-359) x CMSOAZK (1-1201)			
Qy	1	MetAlaThrSer-SerMetLysSerLeuProMetAlaIleProSerPheSerMetCysHi	20
Db	34	ATGGCTACCTCAAGAGCATGAGAGCATTCGAATGGCATCCCAAGTTCTTCATGTGTCA	93
Qy	20	blvleuGlulLeuLeuLysGluGlyLysThrArgAspVal-ProLysAlaGluGluAspG	40
Db	94	CAACCTCGAGCGCTTAAAGAGGAGAACTCGGACACGTGCCAAGCCGAGAGAGATG	153
Qy	40	lulGlyLeuSerCysGluPheGlnGluMetLeuAspSerLeuProLysGluArgLysTPA	60
Db	154	AAGGCTAAGTGGAGTCCAGAGAGATGTGATCTCTTCCTAAGAGAGAGATGGA	213
Qy	60	rgThrArgTyrLeuTyrLeuPheGlnGlyPheTyrCys---GlnAlaLysGluIleGlnA	79
Db	214	GAATCGTACCTTACTTACTTATCCAGAGGTTTGTTGGGGGGGGGGGGGGGTTCGGGG	273
Qy	79	laileMetSerPheGlnLysHisPheGlnSerLeuGluAsnAspValValIleuAlaThrI	99
Db	274	CCATCATGTCTTCTTCCAAACAAATTCCTCAATCCAGTCCGTCTCCGCCACCA	333
Qy	99	leProlLysSerGlyThrThrTyrLeuLysAlaLeuThrPheThrIleLeuAsnArgHisA	119
Db	334	TACCTAAATCCGGTACCACTGGCTGAAAGCTTTAATCTTACATCTTAAACGTCAAC	393
Qy	119	rgPheAspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHisAspL	139
Db	394	GGTTTGATCCGGTGGCTCCGAGTACCAACCACTCTTTCACTTCCAACTCATGACC	453
Qy	139	euValProPhePheGluTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyL	159
Db	454	TTTGATCTTCTTCCGAGTACCAAGCTTTACGCCCAACGAGATGTTCCGATCTTCGGGTC	513
Qy	159	euAlaSerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGluThrIleG	179
Db	514	TAGCAGTCCCAAGACGTTGCCAACCCCACTTACCGTTGGTTCCCTAAAGAGAAAGATCG	573
Qy	179	lulLysProGlyValIleLysValValTyrLeuCysArgAspProPheAspThrPheIleSerS	199
Db	574	AGAAACCCGGTGTAGAGTCTGTGATCTTGTCGGGAACCCGTTGACATTCATCTTT	633
Qy	199	erTPHISerTyrThrAsnAsnLysSerGluSerValSerProValLeuLeuAspGlnA	219
Db	634	CGTGGCATTTACCAACCAATCAATCCGAGTCAAGTCCAGTCTTGTGACCAAG	693
Qy	219	laPheAspLeuTyrCysArgGlyValIleGlyPheGlyProPheThrGluHisIleLeuG	239
Db	694	CTTTGATCTGTATTTGGCGGGAGATGATCGGTTTGGCCCTTTGGGAACAACATGTTGG	753
Qy	239	lyTyrTPArgLysSerLeuLysArgProGluLysValPhePheLeuArgTyrGluAspL	259

Db	754	GATACGTGAGAGAGAGCTTGAGAGACCAAGAAAGTCTTCTTTTAAAGTACGAGATC	813
Qy	259	euLysAspAspIleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuProPheT	279
Db	814	TCAAGACGACATCGAACCACCTTGAGAGGCTTGCAACTTCTTGAAGCTTCTTCA	873
Qy	279	hrgGluGluGluArgLysGlyValValIleLysAlaIleAlaGluLeuCysSerPheGluA	299
Db	874	CCGAGAGAGAGAGAGAGAGAGAGTGTGAAGGTTATCGCGAGCTGTGTAGCTTCGAGA	933
Qy	299	enLeuLysLysLeuGluValAsnLysSerAsnLysSerIleLysAsnPheGluAsnArgP	319
Db	934	ATCTGAAGAAAGTTGGAGTGAACCAAGTCAACCAAGTCAAGTCAAGTCAAGTCAAGTGA	993
Qy	319	heLeuPheArgLysGlyGluValSerAspThrPalaenTyrLeuSerProSerGlnValG	339
Db	994	TCTTGTTCGAAAGAGAGAGAGTGTATGGGTAACATTTGTCACTTCACAAAGTGG	1053
Qy	339	lulArgLeuSerAlaLeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArgLeuS	359
Db	1054	AAGATTTCTACGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1113
Qy	359	er 359	
Db	1114	GC 1115	
RESULT 3			
CMSOAVN			
LOCUS			
DEFINITION			
Arabidopsis thaliana Full-length cDNA Complete sequence from clone			
GSLTPGH42C07 of Hormone Treated Callus of strain col-0 of			
Arabidopsis thaliana (thale cress).			
BX831965			
ACCESSION			
BX831965.1 GI:42458074			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Arabidopsis thaliana (thale cress)			
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;			
rosids; eurosids I; Brassicales; Brassicaceae; Arabidopsis.			
REFERENCE			
1 (bases 1 to 1456)			
Castelli,V., Aury,J.M., Jallion,O., Winkler,P., Clepet,C.,			
Menard,M., Caboche,M., Weissenbach,J. and Salanoubat,M.			
TITLE			
Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences:			
A Combined Approach to Evaluate and Improve Arabidopsis Genome			
Annotation			
JOURNAL			
REFERENCE			
2 (bases 1 to 1456)			
Genoscope.			
AUTHORS			
Direct Submission			
Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :			
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqret@genoscope.cns.fr			
- Web : www.genoscope.cns.fr)			
JOURNAL			
COMMENT			
The sequences are based on single pass reads.			
Life Technologies (a division of Invitrogen) members carried out			
full-length libraries construction : Temple G.			
Genoscope members carried out sequencing and annotation : Castelli			
V., Aury J.M., Jallion O., Winkler P., Menard M., Cruaud C.,			
Schachter V., Weissenbach J., Salanoubat M.			
UNGV INRA : Clepet C., Caboche M.			
Annotation is based on the June 2003 version of the Arabidopsis			
genome released by MIPS (Munich Information center for Protein			
Sequences) . 5 prime and 3 prime are assembled with Piprap.			
http://www.genoscope.cns.fr/externe/sequences/Banque_projec_EF/Full			
length			
http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.			
FEATURES			
source			
1..1456			
/organism="Arabidopsis thaliana"			
/mol_type="mRNA"			
/db_xref="taxon:3702"			
/clone="GSLTPGH42C07"			

/cissue\_type="Hormone Treated Callus"  
/ecotype="Col-0"  
/plasmid="PCMVSPORT\_6"  
complement(1..1456)  
/gene="At5g07010"

## ORIGIN

## Alignment Scores:

Pred. No.: 3,736-193 Length: 1456  
Score: 1778.00 Matches: 337  
Percent Similarity: 96.90% Conservative: 7  
Best Local Similarity: 94.93% Mismatches: 11  
Query Match: 93.68% Indels: 0  
DB: 4 Gaps: 0

US-10-019-931-3 (1-359) x CNS0A0VN (1-1456)

QY 5 SerMetLysSerLysProwerAlaIleProSerPheSerMetCysHisLysLeuGluLeu 24  
Db 73 GCGGTGAAGAGCATTCGAATGGCGATCCCAAGTTTCTCATGTGTACACAGCTCGAGCTC 132  
QY 25 LeuLysGluGlyLysThrArgAspValProlYsAlaGluLysGluLysLeuSerCys 44  
Db 133 CTTAAAGAAAGCGAAACCTCGCAGCTCCGAAAGCGAAAGATGAAGGCTTAAGCTGC 192  
QY 45 GluPheGluGluMetLeuAspSerLeuProlYsGluArgGlyTyrArgThrArgTyrLeu 64  
Db 193 GAGTTCGAAGAGATGTGATCTCTCTTCAAGAGAGAGATGAAGATCTGTTACCTT 252  
QY 65 TyrLeuPheGluGlyPheTyrCysGluAlaLysGluLysLeuAlaIleMetSerPheGlu 84  
Db 253 TACCTATTCCAAAGGTTTGGTGGCCAGCCAAAGAGATTCAAGCATCATGCTTCTCCAA 312  
QY 85 LysHisPheGluInsLeuGluAspValValLeuAlaThrLeuProlYsSerGlyThr 104  
Db 313 AAACATTTCCATCCCTCGATACAGACGCTGCTCGCCACCATACCTAAATCCGTTACA 372  
QY 105 ThrTyrLeuLysAlaLeuThrPheThrIleLeuAsnArgHisArgPheAspProlYsAla 124  
Db 373 ACCTGGCTTAATGCTTAACTTTCACATCTTAAACCGTACACGGTTCATCCGTTGCC 432  
QY 125 SerSerThrAsnHisProlLeuPheThrSerAsnProHisAspLeuValProlPheGlu 144  
Db 433 TCGAATACCAACACCCCTCTTTCATCTTCCAAACCTCATGACCTTGTACCTTCTTCGAG 492  
QY 145 TyrLysLeuTyrAlaAsnGlyAspValProlAspLeuSerGlyLeuAlaSerProlArgThr 164  
Db 493 TACAAAGCTTACGCAACGAGATGTTCCGATCTCTCGGGTCTAGCCAGTCCAAAGAACG 552  
QY 165 PheAlaThrHisLeuProlPheGlySerLeuLysGluThrIleGluLysProlGlyValLys 184  
Db 553 TTCGCAACCCACTTACCGTTCGTTCCCTAAAGAAACATGATTAACCCGTTGTAAG 612  
QY 185 ValValTyrLeuCysArgAsnProPheAspThrPheIleSerSerTyrPheIleThrAsn 204  
Db 613 GTCCGTACTTGTGCGGAAACCGTTTGCACATTCATCTTCTGTCGATTAACCAAC 672  
QY 205 AsnIleLysSerGluSerValSerProlValLeuLysAspGluAlaPheAspLeuTyrCys 224  
Db 673 AACATCAAAATCCGAGTCAGTACCCAGCTTGTGCTAGACCAAGCTTTTATCTGATTGTC 732  
QY 225 ArgGlyValIleGlyPheGlyProlPheTyrGluHisMetLeuGlyTyrTyrArgGluSer 244  
Db 733 CGGGAAGTATGCGGTTTGGCCCGTTTGGAAACATTTGGATATCGAGAGAGAGAC 792  
QY 245 LeuLysArgProlGluLysValPhePheLeuArgTyrGluAspLeuLysAspAspIleGlu 264  
Db 793 TTGAAGAGACCCAGAGATTGTCTTCTTTTACGTACGAGGATCTCAAGACGACATCCGAG 852  
QY 265 ThrAsnLeuLysArgTyrLeuAlaThrPheLeuGluLysProlPheThrGluGluGluArg 284  
Db 853 ACCAATCTGAAGAGGCTGTACTTCTTGTAGAGCTTCTTTCACGAAAGATGAGAAAGCT 912

QY 285 LysGlyValValLysAlaIleAlaGluLeuCysSerPheGluAsnLeuLysLeuGlu 304  
Db 913 GGGGTGTGTGAAGGCTATCGCCGAGCTGTGTAGCTTCGAGAAATCTGAAGAGTTTGAG 972  
QY 305 ValAsnLysSerAsnLysSerIleLysAsnPheGluAsnAlaGluPheLeuPheArgLysGly 324  
Db 973 GTGTACAAAGTCAACACATCATCTAGAACCTTTGAGATCATCTTGTTCGGAATGGA 1032  
QY 325 GluValSerAspTyrPvalAsnTyrLeuSerProlSerGluValGluArgLeuSerAlaLeu 344  
Db 1033 GAACTGACTGTGGGTATCACTATTGTGCACCTTCACAAAGTGAAGATTTGACCTTT 1092  
QY 345 ValAspAspLysLeuGlyGlySerGlyLeuThrPheArgLeuSer 359  
Db 1093 GTGATGACGAGATTAGGTGATCTGTCTCATCTTCTGTGAGC 1137

## RESULT 4

CNS0A186 1517 bp mRNA linear HTC 06-FEB-2004  
LOCUS  
DEFINITION Arabidopsis thaliana Full-length cDNA Complete sequence from clone GSRFPB44ZG02 of Flowers and buds of strain col-0 of Arabidopsis thaliana (thale cress).

ACCESSION BX829897.1 GI:42458543  
VERSION  
KEYWORDS  
SOURCE HTC; GSI, cDNA.  
ORGANISM Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE  
AUTHORS Castelli,V., Aury,J.M., Jallion,O., Wincker,P., Clepet,C., Menard,M., Cruaud,M., Queirer,F., Scarpelli,C., Schachter,V., Temple,G., Caboche,M., Weissenbach J. and Salanoubat M.  
Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences: A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation

TITLE  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1517)  
GENOSCOPE.  
TITLE Direct Submission  
AUTHORS Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)

COMMENT  
Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.  
The sequences are based on single pass reads.  
Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.  
Genoscope members carried out sequencing and annotation : Castelli V., Aury J.M., Jallion O., Wincker P., Menard M., Cruaud C., Schachter V., Weissenbach J., Salanoubat M.  
URGV INRA : Clepet C., Caboche M.

Annotation is based on the June 2003 version of the Arabidopsis genome released by MIPS (Munich Information center for Protein Sequences) . 5 prime and 3 prime are assembled with Phrap.  
http://www.genoscope.cns.fr/externe/sequences/Banque\_Projet\_EF/Full length  
http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.

## FEATURES

source location/Qualifiers  
1..1517  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/clone="GSRFPB44ZG02"  
/cissue\_type="Flowers and buds"  
/ecotype="Col-0"  
/plasmid="PCMVSPORT\_6"  
complement(1..1517)  
/gene="At5g07000"

## ORIGIN

Alignment Scores: 1,466-170 Length: 1517  
Pred. No.: 1582.00 Matches: 305  
Score:

Percent Similarity:	92.46%	Conservative:	26
Best Local Similarity:	85.20%	Mismatches:	21
Query Match:	83.35%	Indels:	6
DB:	4	Gaps:	3
US-10-019-931-3 (1-359) x CNS0A186 (1-1517)			
Qy	4 SerSerMetLysSerIleProMetAlaIleProSerPheSerMetCysHISLysLeuGlu 23		
Db	225 TCAGACATTAAGAGCATTCAGATGCGATCCCAAGTTCTCATTGTGCACAGCCGAG 284		
Qy	24 LeuLeuysgIuGlyLysThrArgAspValProLysalagIuAspGlyLysSer 43		
Db	285 CTCCTTAAGAGAGCAAAAGC-----GAGGCCAAGAAAGAAAGGCGTAAAGC 332		
Qy	44 CysgluPheGluMetLeuAspSerLeuProLysgluArgGlyTPArgThrArgTyr 63		
Db	333 TACGAGTTCAGAGAGATGTTGAGCTCTCTCTTAAGAGAGAGAGCAAGAAATCGTTAC 392		
Qy	64 LeuTyrIleuPheGlnGlyPheTyrCysGlnAlaLysGlnIleGlnAlaIleMetSerPhe 83		
Db	393 CTTACTTATTCAGAGGGGTTCGGTCCAGACCTTAAGAGATTAAGCTATACGCTTTC 452		
Qy	84 GlnLysHISpHeGlnSerLeuGlnAsnAspValValLeuAlaThrIleProLysSerGly 103		
Db	453 CAATAACATTTTCAGCTCCCTCCAGACAGCTGTCTCCGCGACACATACCTTAATCTGCG 512		
Qy	104 ThrThrTyrLeuLysValLeuThrPheThrIleLeuAsnArgHISargPheAspProVal 123		
Db	513 ACAACCTGGTTAAAGCTTTAACTTTCACTTACCATCCCTCAATCGGTGATCCGGTT 572		
Qy	124 ---AlaSerSerThrAsnHISProLeuPheThrSerAsnProHISAspLeuValProPhe 142		
Db	573 TCCCATCAATGTCGAGCACCCCTCTTCTCAATCCAAACCTCAAGACCTGTGATCCCTTC 632		
Qy	143 PheGlnTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAlaSerPro 162		
Db	633 TTTCAGATCAAGACTTTACGCCCAAGGAAATGTTCCCATCTCTCGGGCTACGCCAGTCCA 692		
Qy	163 ArgThrPheAlaThrHISLeuProPheGlySerLeuLysGlnThrIleGlnLysProGly 182		
Db	693 AGAACATTCGCAACCCAGTACCGTTCGGTCCCTTAAGATTCGGTGAAGAAATCCCACT 752		
Qy	183 ValLysValValTyrLeuCysArgAsnProPheAspThrPheIleSerSerThrHISLys 202		
Db	753 GTGAAGGTGTGATCCTGTGCGGAAACCGTTTGAACATTCATCTCCATGTGGCATTAAC 812		
Qy	203 ThrAsnAsnIleLysSerGlnSerValSerProValLeuLeuAspGlnAlaPheAspLeu 222		
Db	813 ATCAACAACATCACTTCGCAATCGATGAGCCAGCTTTGCTAGACGAAGCTTTTGATCTA 872		
Qy	223 TyrCysArgGly---ValIleGlyPheGlyProPheTyrGlnHISmetLeuGlyTyrTyr 241		
Db	873 TATGTCCGGGGATTAACGATCGGATTTGGCCGTTTGGAAACATATGTGGATCTGCG 932		
Qy	242 ArgGlnSerLeuLysArgProGlnLysValPhePheLeuArgTyrGlnAspLeuLysAsp 261		
Db	933 AGAGAGAGCTTGAAGAGCGCAGAGAAAGCTTATTTTAAAGTACGAGAGATCCAAAGAA 992		
Qy	262 AspIleGlnThrAsnLeuLysArgLeuAlaThrPheLeuGlnLeuProPheThrGlnGlu 281		
Db	993 GACATTCAGACCAACTTCAGAAAGACTAGCAAGTTCTTTAGACTTCCTTTACCGAAGAA 1052		
Qy	282 GlnGlnArgLysGlyValValLysAlaIleAlaGlnLeuCysSerPheGlnAsnLeuLys 301		
Db	1053 GAGGAACAAAGAGGAGTGTGAAGCTATCGCTGATCTGTGTAGCTTTGAGAATCTGAAG 1112		
Qy	302 LysLeuGlnValLeuLysSerAsnLysSerIleLysAsnPheGlnAsnArgPheLeuPhe 321		
Db	1113 AAGTTGAGGAGTGAACAGTCAAGCAATTTGATCCGAACATATGAGAACCGGTTCTTGTTT 1172		
Qy	322 ArgLysGlnGlnValLeuSerAspTyrValAsnTyrLeuSerProSerGlnValGlnArgLeu 341		

Db	1173 AGAAGAGAGAGATGATGATTGTTACTTATTTGCGCATCAAGTGAAGAAATTTG 1232		
Qy	342 SerLysLeuValAspAspLysLeuGlyLysSerGlyLeuThrPheArgLeuSer 359		
Db	1233 TCAGCCTTAAGTGAATGACAAAGTTAGCTGATCTGTCTCACTTTCAGATTTGAGT 1286		
RESULT 5			
CNS0A1K2			
LOCUS			
DEFINITION			
CNS0A1K2 1320 bp mRNA linear HTC 06-FEB-2004			
Arabidopsis thaliana Full-length cDNA Complete sequence from clone			
GSLTFRB142D09 of Flowers and buds of strain col-0 of Arabidopsis			
thaliana (thale cress).			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segret@genoscope.cns.fr - Web : www.genoscope.cns.fr)			
The sequences are based on single pass reads.			
Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.			
Genoscope members carried out sequencing and annotation : Castelli V., Aury J.M., Jailon O., Wincker P., Menard M., Cruaud C., Schachter V., Weissenbach J., Salanoubat M.			
UNGV INRA : Clapet C., Caboche M.			
Annotation is based on the June 2003 version of the Arabidopsis genome released by MIPS (Munich Information center for Protein Sequences). 5 prime and 3 prime are assembled with Phrap.			
http://www.genoscope.cns.fr/externe/sequences/Barque_Projet_EF/Full_length			
http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.1.1320			
Location/Qualifiers			
1.1320			
/organism="Arabidopsis thaliana"			
/mol_type="mRNA"			
/db_xref="taxon:3702"			
/clone="GSLTFRB142D09"			
/tissue_type="Flowers and buds"			
/ecotype="Col-0"			
/plasmid="pCMVSPORT_6"			
complement (1.1320)			
/gene="At5g07000"			
ORIGIN			
Alignment Scores:			
Pred. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			
DB:			
US-10-019-931-3 (1-359) x CNS0A1K2 (1-1320)			
Qy	4 SerSerMetLysSerIleProMetAlaIleProSerPheSerMetCysHISLysLeuGlu 23		
Db	29 TCAGACATTAAGAGCATTCAGATGCGATCCCAAGTTTCTCATTGTGCACAGCCGAG 88		

```

QY      24 LeuLeuLysGluGlyLysThrArgAspValProLysAlaGluGluAspGluGlyLeuSer 43
DB      89 CTCCTTAAGGAAGGAAAAGC-----GAAAGCCAGAAAGAAAGGCTTAAGC 136
QY      44 CysGluPheGlnGluMetLeuAspSerLeuProLysGluArgGlyTTPArgThrArgTyr 63
DB      137 TACGAGTTCCAAAGAGATGTTGCACTCTCTTAAGAGAGAGACGGGAATGCTTAC 196
QY      64 LeuTyrLeuPheGlnGlyPheTTPCysGlnAlaLysGluIleGlnAlaIleMetSerPhe 83
DB      197 CTTTACTTAATCCAAAGGGTTTCGGTCCAGCTTAAGAGATTCAAAGCTTACGCTTTC 256
QY      84 GlnLysHisPheGlnSerLeuGluAspValValLeuAlaThrIleProLysSerGly 103
DB      257 CAAAACATTTTCACTCCCTTCCAGACGAGTGTCTCTGCGCACCATCTTAATCTGGC 316
QY      104 ThrThrTPLeuLysAlaLeuThrPheThrIleLeuAsnHisArgPheAspProVal 123
DB      317 ACAACCTGGTTAAAGCTTTAACTTTACCATCCCTTACCCTGATCGGTTGATCCGGTT 376
QY      124 ---AlaSerSerThrAsnHisPheLeuPheThrSerAsnProHisAspLeuValProPhe 142
DB      377 TCCTATCAAGTTCCGACCACTCTCTCTCACATCCACCTTCAAGACCTCGTACCTTTC 436
QY      143 PheGluTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAlaSerPro 162
DB      437 TTCGAGTCAAGACTTTACGCCCAAGGAAATGTTCCGAGTCTCGGGTCTACCGAGTCCA 496
QY      163 ArgThrPheAlaThrHisLeuProPheGlySerLeuLysGluThrIleGluLysProGly 182
DB      497 AGAATATTTCGCAACCAACGATACCGTTCGGTGCCTTAAGATTCGATCGACGAATCCCACT 556
QY      183 ValLysValIValTyrLeuCysArgAsnProPheAspThrPheIleSerSerTrpHisTyr 202
DB      557 GTGAAGTTGTGTACTCTGTGCGGAAACCGCTTTGACACTTATCTCCATGTGGCATTC 616
QY      203 ThrAsnAsnIleLysSerGluSerValSerProValLeuLeuAspGlnAlaPheAspLeu 222
DB      617 ATCAACAACATCATCTTCGAGTCAAGTCAAGCGAGCTTCTGACGAAGACTTTGATCTGA 676
QY      223 TyrCysArgGly---ValIleGlyPheGlyProPheTrpGluHisMetLeuGlyTyrTrp 241
DB      677 TATTCCTCGGGGATTACTGATTCGATTTGGCCCGTTCGGAACACATGTGGGATATCGG 736
QY      242 ArgGluSerLeuLysArgProGluLysValPhePheLeuArgTyrGluAspLeuAsp 261
DB      737 AGAGAGAGCTTGAAGAGCCAGAGAAAGTCTTAATTTTAAAGTACGAGGATCTCAAGAGA 796
QY      262 AspIleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuProPheThrGluGlu 281
DB      797 GACATCGAGACCACTTGAAAGAGCTAGCAAGATTCTTAGGACTTCCTTCCGACGAGAA 856
QY      282 GluGluLysGlyLysValValLysAlaIleAlaGluLeuCysSerPheGluAsnLeuLys 301
DB      857 GAGGAACAAAAGAGGAGTGTGAAGCTATCGCTGATCTGTGTAGCTTTGAGAAATCTGAAG 916
QY      302 LysLeuGluValAsnLysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeuPhe 321
DB      917 AAGTTGAGAGCTGAACAGTCAAGCAATTTGATCCAGACTATGAGAACCGGTTCTTGT 976
QY      322 ArgLysGlyGluValSerAspTrpValAsnTyrLeuSerProSerGlnValGluArgLeu 341
DB      977 AGGAAGAGGAGAGTGTGATTTGGTTAACTATTGTGCGCATCAAGATGGAAAGATTG 1036
QY      342 SerAlaLeuValAspAspLysLeuGlyLysSerGlyLeuThrPheArgLeuSer 359
DB      1037 TCAGCCTTAGTGATGACAAAGTTAGCTGATCTGTCTCACTTTCAGATTGAGT 1090

```

```

RESULT 6
CNS09YST 1302 bp mRNA linear HTC 06-FEB-2004
LOCUS Arabidopsis thaliana Full-length cDNA Complete sequence from clone
DEFINITION

```

```

ACCESSION BX831024.1 GI:42454963
VERSION Arabidopsis thaliana (thale cress).
KEYWORDS HRC; GSTL cDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE Castell, V., Aury, J.M., Jallou, O., Wincker, P., Clepet, C., Menard, M., Cruaud, C., Quetier, F., Searpelli, C., Schachter, V., Temple, G., Caboche, M., Weissenbach, J., and Salanoubat, M. Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences: A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation
TITLE Unpublished
AUTHORS 2 (bases 1 to 1302)
REFERENCE Genoscope.
JOURNAL Direct Submission
TITLE Submitted (18-NOV-2003) Genoscope - Centre National de Sequence : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
JOURNML - Web : www.genoscope.cns.fr)
COMMENT The sequences are based on single pass reads. Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G. Genoscope members carried out sequencing and annotation : Castell V., Aury J.M., Jallou O., Wincker P., Menard M., Cruaud C., Schachter V., Weissenbach J., Salanoubat M.
FEATURES
source
location/Qualifiers
1..1302
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/cisue="GSTLS502A11"
/cisue_type="Adult vegetative tissue"
/ecotype="Col-0"
/plasmid="PCMTSPORT 6"
complement(1..1302)
/gene="At5g07000"
ORIGIN
Alignment Scores:
Pred. No.: 5.09e-165 Length: 1302
Score: 1533.00 Matches: 296
Percent Similarity: 91.04% Conservative: 29
Best Local Similarity: 82.91% Mismatches: 26
Query Match: 80.77% Indels: 6
DB: 4 Gaps: 3
US-10-019-931-3 (1-359) x CNS09YST (1-1302)
QY      4 SerSerMetLysSerIleProMetAlaIleProSerPheSerMetCysHisLysLeuGlu 23
DB      12 TCAAGATCAAAAGAGATTTCCATGCGGATCCCAAGTTTCTCATGTGTCAACAAGCCGAG 71
QY      24 LeuLeuLysGluGlyLysThrArgAspValProLysAlaGluGluAspGluGlyLeuSer 43
DB      72 CTCCTTAAGGAAGGAAAAGC-----GAAAGCCAGAAAGAAAGGCTTAAGC 119
QY      44 CysGluPheGlnGluMetLeuAspSerLeuProLysGluArgGlyTTPArgThrArgTyr 63
DB      120 TACGAGTTCCAAAGAGATGTTGCACTCTCTTAAGAGAGAGACGGGAATGCTTAC 179
QY      64 LeuTyrLeuPheGlnGlyPheTTPCysGlnAlaLysGluIleGlnAlaIleMetSerPhe 83

```

```

Db      ||||| 180 CTTACTTATTCCAAAGGTTCCGGTCCCAAGCTAAGAGATTCACGATTCACGCTTTTC 239
Qy      ||||| 84 GlnlyshispeginserleuGluasnApyValleuAathrileProlySergly 103
Db      ||||| 240 CAAAACATTTTCAGTCCCTTCACAGACGATGCTCCGCAACATCTAAATCTGGC 299
Qy      ||||| 104 ThrThrtripleuysAlaleuThrPheThrileleuAanrghisArgPhasPProVal 123
Db      ||||| 300 ACAACCTGGTTAAAGCTTTAACTTCACTTCAACCTTCACCTTCATCGGTTGATCCGGTT 359
Qy      ||||| 124 ---AlaserSerThrAenhiProleuPhePheSerAenProhisAapLeuValProPhe 142
Db      ||||| 360 TCCCATCAATGTTCCGACCACTCTTCTTCACATCCAACTTCACGACCTGATCCTTTC 419
Qy      ||||| 143 PheGlnTyrlsleuTyrlAasnGlyAspValProApleuSerglyLeuAlaserPro 162
Db      ||||| 420 TTCAGATCAAGCTTTAAGCCCAAGAAATGTTCCGATCTCTCGGCTCTAGCCAGTCCA 479
Qy      ||||| 163 ArgThrPheAlatrhrHisleuProPheGlySerglySerglyuThrileGlnlyProGly 182
Db      ||||| 480 AGAACATTCGCAACCCAGTACCGTTCGTCCTTAAAGATTCGGTCAAGAAATCCCGCT 539
Qy      ||||| 183 VallyserValTyrlsleuCyArgAanProPheAapThrPheileSerSerTriphistyr 202
Db      ||||| 540 TTTAAGGTGTTTACTCTGCGGACCCGTTTACACATTCATCTCCATGTTGATTC 599
Qy      ||||| 203 ThrAanenielySerglySerglySerglyValleuAapGlnAlaPheAspleu 222
Db      ||||| 600 ATCAACAACATCACTTCGAGTCAAGTGAAGACATCTTGAAGACGATTTGATCTA 659
Qy      ||||| 223 TyrCyAsargly---ValileGlyPheGlyProPheTrpGlnHisMeleuGlyTyrTrp 241
Db      ||||| 660 TATTCGCGGATTAATCTGATCGGATTTGCGCGTTTGCGAACAATGTTGGATACG 719
Qy      ||||| 242 ArgGlyserleuLyArgProGlyLyValPhePheleuAarglyGlyAspleuLyAsp 261
Db      ||||| 720 AGAAGAGCTTGAAAGAGCCAGAAAGTCTTATTTTAAAGTACGAGATCTCAAAAGA 779
Qy      ||||| 262 AspIleGlnThrAenleuLyArgleuAlaThrPheleuGlnleuProPheTrpGln 281
Db      ||||| 780 GACATCGAGACCAACTTGAAAGAGCAAGATTTCTTGAAGACTTCCTTCAACGAGAA 839
Qy      ||||| 282 GlnGlnarglySerglyValleuAlaIleleuGlnleuCySergPheGlnleuLyS 301
Db      ||||| 840 GAGAAACAAAGGAGGTTGTGAAGCTATCGCTGATGCTGATGAGATCTGAAG 899
Qy      ||||| 302 LysleuGlnValAenLySerglySerglySerglyAanPheGlnAanArgPheleuPhe 321
Db      ||||| 900 AAGTTGAGGTGAACACTCAAGCAAAATTCATCCAACTATGAGAACCGGTTCTTGT 959
Qy      ||||| 322 ArglySerglyGlnValSergAspTrpValAanTyrlsleuSergProSerglyValGlnArgleu 341
Db      ||||| 960 AGGAAAGAGAGAGTCACTTGTGTTAATCATCTGTCGCATTCACAGTGAAGAAAGATTG 1019
Qy      ||||| 342 SerAlaIleuValAapAspLySerglySerglySerglyLeuThrPheArgleu 358
Db      ||||| 1020 TCAGCTTAAGTGAAGTCAAGTACGATCTGATCTCACCCTTCAAGATT 1070

RESULT 7
CNS0A0WP 1328 bp mRNA linear HTC 06-FEB-2004
LOCUS DEFINITION
Arabidopsis thaliana full-length cDNA complete sequence from clone
GSLTPGH36ZF09 of Hormone Treated Callus of strain col-0 of
Arabidopsis thaliana (thale cress).
ACCESSION
BX831880
VERSION
BX831880.1 GI:42458032
KEYWORDS
HTC; GST; cDNA
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

```

```

REFERENCE 1 (bases 1 to 1328)
AUTHORS Castell, V., Aury J.M., Jallion O., Wincker P., Clepet C.,
Menard M., Cruaud C., Quetier F., Scareselli C., Schachter V.,
Temple G., Caboche M., Weissenbach J. and Salanoubat M.
TITLE Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
A Combined Approach to Evaluate and Improve Arabidopsis Genome
Annotation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1328)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out
full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castell
V., Aury J.M., Jallion O., Wincker P., Menard M., Cruaud C.,
Schachter V., Weissenbach J., Salanoubat M.
UNREV INRA : Clepet C., Caboche M.
Annotation is based on the June 2003 version of the Arabidopsis
genome released by MIPS (Munich Information center for Protein
Sequences) . 5 prime and 3 prime are assembled with Phrap.
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
length
http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
FEATURES
source Location/Qualifiers
1..1328
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="GSLTPGH36ZF09"
/risue_type="Hormone Treated Callus"
/ecotype="Col-0"
/plasmid="PCWVSPORT 6"
complement(1..1328)
/gene="At5g07010"
gene
ORIGIN
Alignment Scores:
Pred. No.: 1,97e-164 Length: 1328
Score: 1528.00 Matches: 314
Percent Similarity: 89.97% Conservative: 9
Best Local Similarity: 87.47% Mismatches: 25
Query Match: 80.51% Indels: 11
DB: Gaps: 2
US-10-019-931-3 (1-359) x CNS0A0WP (1-1328)
Qy 8 SerIleProMet-AlaIleProSerPhe-SerMetCyHisIleuGlnleuLeuLySg 27
Db 65 AGCATTCCAATGGGGGGTCCCAAGTTCTTTTTCATCAATCTCCACCTCTTAAAG 124
Qy 27 LnglyLySgThr---ArgAspValProLySAlaGlnGlyAspGlyGlyLySergCySgUp 46
Db 125 AAGCAAAACTCCCCACCGCTCCCGCAAAACCCCAACAATGAAGGCTAAGTCCAGT 184
Qy 46 heGlnIleuMeleuAapSergleuProLySgIuArgGlyTTPArgThrArgTyrl 66
Db 185 TCCAAAGACATGTTGATCTCTTCTTAAGAGAGAGATGGAAGTACCTTACC 244
Qy 66 eupheGlnGlyPheTrpCySgIlnAlaLySgIleGlnAlaIleMetSergheGlnYsh 86
Db 245 TATTCCAAGGGTTTGTGTCACCAAGGATTCAGAGCCATCATGCTTTCCAAAGGG 304
Qy 86 isPheGlnSergleuGlnAanAspValValleuAlaThrileProLySerglyThrTrp 106
Db 305 ATTTCCAATCCCTCGAAGACAGCTCTTCTCGCACCACTTAATGCGTCAACT 364
Qy 106 rpleuLyAlaIleuThrPheThrileleuAanrghisArgPheAspProValAlaserS 126
Db 365 GGCTAAGGCTTTAATCTTCAACATCTTAACTCAACCGGTTGATCCGGTTCCTCGA 424

```



QY 126 eTThrAsnHisProLeuPheThrSerAsnProHisAspLeuValProPhePheGluTyrL 146  
 DB 425 GTCACCAACCACTCTTTCACTTCCAACTCCTAGACCTTGTACTCTTCTCGAGTACA 484  
 QY 146 yAlLeuTyrLaangIyAspValProAspLeuSerGlyLeuLAspProArgThrPheA 166  
 DB 485 ACCTTACCGCCAAACGAGATGTTCCGATCTCTCGGATCTAGCCAGTGCAGAACGCTTCG 544  
 QY 166 lATHisLeuProPheGlySerLeuValGluThrLLeuLysProGlyValLysValV 186  
 DB 545 CAACCCACTTACCGTTCGGTCCCTTAAAGAAACGATTCAGAAACCCGGTGTGAAGTGC 604  
 QY 166 aLYLeuCyAspAsnProPheAspThrPheLLeuSerSerTrpHisTyrThrAsnAsn 206  
 DB 605 TGTACTCTGCGCGAACCCTTTCAAACTTCATCTCTTCGTGGCATTCACCAACACACA 664  
 QY 206 lElysSerGluSerValSerProValLeuLeuAspGlnLaPheAspLeuTyrCyAspArg 226  
 DB 665 TCAATCCGAGTGAAGTGAAGCCAGTCTTCTAGACCACTTTGATCTGTATGCGCGG 724  
 QY 226 lYValIleGlyPheGlyProPheTrpGluHisMetLeuGlyTyrTrpArgGluSerLeuL 246  
 DB 725 GAGTATCGGGTGTGGCCGTTTGGAAACACATGTTGGAAATCTGGAGAGAGAGCTTGA 784  
 QY 246 yARgProGluLysValPhePheLeuArgTyrGluAspLeuLysAspAspIleGluThrA 266  
 DB 785 AAGAGCCAGAGAAAGTCTTCTCTTAGGTACGAGGCTTCAAAAGACGACATCGAGACA 844  
 QY 266 eNLeuLysArgLeuValaThrPheLeuGluLeuProPheThrGluGluGluValArg 286  
 DB 845 ACTTGAAAGAGGCTTGCACTTCTTAGAGCTCCCTTCCACGAGAAAGGAGAAAGG 904  
 QY 286 lYValValLysValaIleAlaGluLeuCySerPheGluAsn--LeuLysValLeuGluValA 305  
 DB 905 GAGTGTTAAGGCTATCGCCGAGCTGTGAGCTTCGAGAACTCTCGAAGAAAGTGGAGT 964  
 QY 305 l-AspLysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeu--PheArgLysG 324  
 DB 965 CGAACAGTGAAGAACAGTGAATCAAGAACTTTGAGATGATTCCTTGTTGTTGGAAAGG 1024  
 QY 324 yGluValSerAspTrpValAsnTyrLeu--SerProSerGluValGluArgLeuSerAla 343  
 DB 1025 ACAAAGTGAATGATGGGTTCACTATTGTTCACACCTTCACAAGTGAAGATGTTCAGCC 1084  
 QY 344 LeuValAspAsp--LysLeuGlyGlySerGlyLeuThr 355  
 DB 1085 TTAGTGAATGACCAAGCTTAGGTGATCTGTTCTCACT 1123  
 RESULT 8  
 LOCUS BZ062934/c 757 bp DNA linear GSS 10-OCT-2002  
 DEFINITION 11c02b11.b1 B.oleracea002 Brassica oleracea genomic, genomic survey  
 sequence.  
 ACCESSION BZ062934  
 VERSION BZ062934.1 GI:23673253  
 KEYWORDS GSS.  
 SOURCE Brassica oleracea  
 ORGANISM Brassica oleracea  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosoids; eurosids II; Brassicales; Brassicaceae; Brassica.  
 REFERENCE 1 (bases 1 to 757)  
 AUTHORS Delahunty,K., Fewell,G., Pullon,L., McCombie,W.R., Miner,T.,  
 Naeh,W., Rabinowicz,P.D. and Wilson,R.K.  
 TITLE Whole genome shotgun reads from Brassica oleracea  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Richard K. Wilson  
 Genome Sequencing Center  
 Washington University School of Medicine  
 Email: submis@wustl.wustl.edu  
 Plate: 11c02 row: b column: 11  
 Seq primer: -21UPOT forward

Class: shotgun  
 High quality sequence start: 52  
 High quality sequence stop: 551.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..757  
 /organism="Brassica oleracea"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3712"  
 /clone\_lib="B.oleracea002"  
 /note="Vector: pOTw13; Whole genome shotgun library from  
 flowering buds. DNA was purified from a crude nuclear  
 prep using Brassica oleracea 101000D3 buds provided by  
 Thomas Osborn at the University of Wisconsin. Genomic  
 DNA was provided by Pablo Rabinowicz (CSHL) and the  
 shotgun library prepared at Washington University Genome  
 Sequencing Center."  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6.39e-116 Length: 757  
 Score: 1103.50 Matches: 201  
 Percent Similarity: 92.53% Conservative: 22  
 Best Local Similarity: 83.40% Mismatches: 15  
 Query Match: 58.14% Indels: 3  
 DB: 9 Gaps: 1  
 US-10-019-931-3 (1-359) x BZ062934 (1-757)  
 QY 38 GluAspGluGlyLeuSerCyAspGluPheGlnLwMetLeuAspSerLeuProLysGluArg 57  
 DB 755 GAAAGAAAGAGGCTTAAGTCAAGTCTTAAAGATGTTGAGTCTCTTCTTAAGAGAGA 696  
 QY 58 GLYTPARGTARGTARGTyrLeuTyrLeuPheGlnGlyPheTrpCyGlnAlaLysGluL 77  
 DB 655 GCTTGAGAAACTCGTCATCTTACTCTGTTCCAAGGCTTTTGTCACCAAGCAAGAGATT 636  
 QY 78 GlnAlaIleMetSerPheGlnLysHisPheGlnSerLeuGluAsnAspValaLeuAla 97  
 DB 635 CATGCCATCATGTCCTTCCAGAAACATTTCAAGCCTTCCAAAGAGATATCTCTCT 576  
 QY 98 ThrIleProLysSerGlyThrThrTyrLeuLysAlaLeuThrPheThrIleLeuAsnArg 117  
 DB 575 ACCATACCAAAATCCGGTACAAATGTTAAAGCTTCAACATTTACCTCTTAACCGA 516  
 QY 118 HisArgPheAspProValaLysSerSerThrAsnHisProLeuPheThrSerAsnProHis 137  
 DB 515 CACCGGTTGATCCGGT-----TCTGACCACTCTCTCCACATCAAAACCTCAT 465  
 QY 138 AspLeuValProPhePheGluTyrLysLeuTyrAlaAngIyAspValProAspLeuSer 157  
 DB 464 GACCTCGTACCTTCTTGTGATACAGCTTACCGCAACGAGAAAGTTCTGATCTCTCC 405  
 QY 158 GlyLeuAlaSerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGluThr 177  
 DB 404 AGCTTAGCTAGTCCAAAGAAATTCGCAACACACGCTCTTTCGTTCTCTCAAGGCTCA 345  
 QY 178 lLeGluLysProGlyValaLysValaValaTyrLeuCyAspAsnProPheAspThrPheIle 197  
 DB 344 ATCAGAGAACCGGAGTGAAGAGTGTGTGATCTGTGTGCGAACCCGTTGACACGTTCTC 285  
 QY 198 SerSerTrpHisTyrThrAsnAsnLysSerGluSerValSerProValLeuLeuAsp 217  
 DB 284 TCTCTGTGCACTTTCAGACACGATTAAGTCGAGTGGTGAATCCAGTCTCGTTGGAA 225  
 QY 218 GlnAlaPheAspLeuTyrCyAspArgLysValaIleGlyPheGlyProPheTrpGluHisMet 237  
 DB 224 GAAGGTTTGAATCTGTATTCAGAGGAGTGAATCGGGTTCGAGCCGTTTGGAAACACATG 165  
 QY 238 LeuGlyTyrTPARGLysSerLeuLysArgProGluLysValaPhePheLeuArgTyrGlu 257  
 DB 164 CTGGGATCTGAGAGAGAGAGCTTGAAGACCAAGAGAAAGTCTTCTTAAAGTACAA 105  
 QY 258 AspLeuLysAspAspIleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuPro 277



```

Db      104 GATGTAAGAGATATTGACCTAATTGTAAGAGCTTGCAAGTTCTTGCGGCTTCT 45
Qy      278 Phe 278
Db      44 TTC 42

RESULT 9
LOCUS   BH733391
DEFINITION BOMITUOTR BO 2.3 KB Brassica oleracea genomic clone BOMITU60,
ACCESSION BH733391
VERSION   BH733391.1 GI:18838786
KEYWORDS  GSS.
SOURCE    Brassica oleracea
ORGANISM  Brassica oleracea
REFERENCE 1 (bases 1 to 811)
AUTHORS   Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
           Utechtack,T.R., Wortman,J.R., White,O.R. and Town,C.D.
           Whole genome shotgun sequencing of Brassica oleracea and its
           application to gene discovery and annotation in Arabidopsis
           Genome Res. 15 (4), 487-495 (2005)
JOURNAL   PUBMED
COMMENT   15805490
           Other GSSs: BOMITU60TF
           Contact: Chris Town
           TIGR
           9712 Medical Center Drive, Rockville, MD 20850, USA.
           Tel: 301-838-3523
           Fax: 301-838-0208
           Email: cdtown@tigr.org
           DNA is from a doubled haploid provided by Tom Osborn.
           Seq primer: TR
           Class: sheared ends.
FEATURES
    source
        1..811
        /organism="Brassica oleracea"
        /mol_type="genomic DNA"
        /strain="TO1000DH3"
        /db_xref="taxon:3712"
        /clone="BOMITU60"
        /clone_lib="BO 2.3 kb"
        /note="Vector: pHOSt1; Site 1: BstXI; 2-3 kb sheared
        genomic DNA inserted into pHOSt1 using BstXI linkers"
ORIGIN
Alignment Scores:
Pred. No.:      8.79e-115      Length:      811
Score:          1094.00      Matches:      199
Percent Similarity: 93.72%      Conservative: 25
Best local Similarity: 83.26%      Mismatches: 15
Query Match:    57.64%      Indels:      0
DB:             9      Gaps:      0
US-10-019-931-3 (1-359) x BH733391 (1-811)

Qy      119 ArgGheAPpProValAlaSerSerThrAsnHisProLeuPheThrSerAspProHisasp 138
Db      3 CGGTTGATTCGGGTTTCTTCAACCAACCAACCACTCCATCAATCAAAACCTCATATAC 62

Qy      139 LeuValProPhePheGluTyrTylLeuTyrAlaAsnGlyaspValProAspLeuSerGly 158
Db      63 CTCGTAACCTTTCTTGATCAAGCTTTACGAGACGGGATGTTCTGATCTCCCGGT 122

Qy      159 LeuAlaSerProAlaGlnPheAlaThrHisLeuPProPheGlySerLeuLeuGluThrIle 178
Db      123 TTAGCAGACTCGAGAACATTCGCAACCAACGCTCTTCGTTCCCTCAAGAGACTCCATC 182

Qy      179 GluLysPProGlyValIleValValTyrLeuGysArgAsnProPheAspThrPheIleSer 198

```

```

Db      183 GAGGAACACGACGCGAAGTGTGTACTTTGCCGAACACCGGTTGACACATTCATCTCC 242
Qy      199 SerTriPHisTyrThrAsnAniIleYsSerGluSerValSerProValLeuLeuAspGln 218
Db      243 TCGTGCAATTACATCAACCAATMAAGTCGAGGACATGATGCCGTCCTGTAGAAAGAA 302

Qy      219 AlaPheAspLeuTyrCysArgGlyValIleGlyPheGlyProPheTProGluHisMetLeu 238
Db      303 GGGTTTATCTGATCTGATCCGAGGGGTGATCGGCTCGACCGTTTGGAAACATGTTG 362

Qy      239 GlyTyrTriPArgGluSerLeuLysArgProGluLysValPhePheLysArgTyrGluAsp 258
Db      363 GGATACTGGAAGAAGACCTTGAGACACACAGAAAGCTTGTTCTTACGTTAGAAAGAT 422

Qy      259 LeuLysAspAspIleGluTyrThrAsnLeuLysArgGluAlaThrPheLeuGluLeuProPhe 278
Db      423 CTGAACAAAGACAGAGGCTTAACTTAAAGAAAGCTTCAAGCTTCTTGAGAGTTCTTC 482

Qy      279 ThrGluGluGluGluArgLysGlyValValIleValAlaIleAlaGluLeuCysSerPheGlu 298
Db      483 ACGGAAGAGAGGAAACAAAAGGTGTGTGTCATTCATTCGATCTGTGACGCTTGAG 542

Qy      299 AsnLeuLysIleLeuGluValAlaAsnLysSerAsnLysSerIleLysAsnPheGluAsnArg 318
Db      543 AATCTAAGAAGACTGAGAGTGAAACAAGTCAAAATCGATCAAGAACTTTGAGAAATAGG 602

Qy      319 PheLeuPheArgLysGlyGlyValSerAspTyrValAsnTyrLeuSerProSerGlnVal 338
Db      603 CACTGTTCAGAAAGAGAGAGTAGATGATGGTTAATCATCTCTCCCAACCAAGGCC 662

Qy      339 GluArgLeuSerAlaLeuValAlaAspAspLeuGluGlySerGlyLeuThrPheArg 357
Db      663 GAAAGATTTCAGGCTTATGATGAAATGACAGTTAGGTGTTATGCTTCATCTTCAGG 719

RESULT 10
LOCUS   CD837475
DEFINITION BN45.052G01F020104 BN45 Brassica napus cDNA clone BN45052G01, mRNA
ACCESSION CD837475
VERSION   CD837475.1 GI:32519415
KEYWORDS  EST.
SOURCE    Brassica napus (rape)
ORGANISM  Brassica napus
REFERENCE 1 (bases 1 to 727)
AUTHORS   Genopiante, a major partnership french program in plant genomics
TITLE     Unpublished (2003)
JOURNAL   Genopiante
COMMENT   Contact: Genopiante
          93, rue Henri Rochefort 91025 EVRY CEDEX France
          Tel: 33 1 69 47 54 00
          Fax: 33 1 69 47 54 10
          This sequence has been generated in the framework of the french
          plant genomics programme 'Genopiante' (http://www.genopiante.com
          and http://genopiante-info.infobiogen.fr).
FEATURES
    source
        1..727
        /organism="Brassica napus"
        /mol_type="mRNA"
        /culturivar="Jet Neuf"
        /db_xref="taxon:3708"
        /clone="BN45052G01"
        /issue_type="seed"
        /clone_lib="BN45"
ORIGIN
Alignment Scores:
Pred. No.:      2.4e-100      Length:      727
Score:          968.00      Matches:      186

```

Percent Similarity: 90.39% Conservative: 21  
 Best Local Similarity: 81.22% Mismatches: 14  
 Query Match: 51.00% Indels: 8  
 DB: 6 Gaps: 5

US-10-019-931-3 (1-359) x CD834047 (1-727)

```

QY 1 MetAlaThrSerSerMetLySerIlePro---MetAlaIleProSerPheSerMetCys 19
DB 51 ATGGGACCTCAAGACATCAAGAGTCTTCATTAATGCGATCCCAAGTTCTCCATTGCG 110
QY 20 HisLysLeuGluLeuLeuLys---GluGlyLysThrArgAspValProLys-----Ala 36
DB 111 CACAAGCAGAGGCTCTCAAGAGAGAGCCAAAGCCGAGAC---CCGAAAGCCCAAGAA 167
QY 37 GluGluAspGluGlyLeuSerCysGluPheGlnGluMetLeuAspSerLeuProLysGlu 56
DB 168 GAAGAAGAAAGAGGGCTTAAGCTTAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 227
QY 57 ArgGlyTPArgThrArgTyrLeuTyrLeuPheGlnGlyPheTyrCysGlnAlaLysGlu 76
DB 228 AGAGTTGGAGAACCTCGTCATCTTTACCTGTTCCAAAGCTTTTGGTGGCCAAAGCAGAG 287
QY 77 IleGlnAlaIleMetSerPheGlnLysHisPheGlnSerLeuGluAsnAspValValLeu 96
DB 288 ATTCATGCCATCATGTCTTCCAGAAACATTTCAAGACTCTCCCAAAAGATGTATTCTC 347
QY 97 AlaThrIleProLysSerGlyThrThrTrpLeuLysAlaLeuThrPheThrIleLeuAsn 116
DB 348 GGCACCATTAACCAAAATCCGGTACCAATGGTTAAAGCTTTAAACCTTAACCTCTTAAC 407
QY 117 ArgHisArgPheAspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnPro 136
DB 408 CGACACCCGTTTGAATCCGTT-----TCGACCACTCTCTCTCACAACAACCTT 458
QY 137 HisAspLeuValProPhePheGluTyrLysLeuTyrAlaAsnGlyAspValProAspLeu 156
DB 459 CATGACCTCGTACTTTCTTGGAGTACAAAGCTTTACGCCAAGAGAGTTCTGATCTC 518
QY 157 SerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGlu 176
DB 519 TCCGGTCTAGCTAGTTCCAAGAACATTCGCAACACAGCTCTTCCGTTCTCTCAAGGGC 578
QY 177 ThrIleGluLysProGlyValLysValValTyrLeuCysArgAsnProPheAspThrPhe 196
DB 579 TCCATCGAAGAACCGGAGCGAAGGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 638
QY 197 IleSerSerTPHisTyrThrAsnAsnIleLysSerGlySerValSerProValLeuLeu 216
DB 639 ATCTCTCTGTGCAATTCAGCAACAGCATTAAGTGGAGTCCGTTGAGTCCAGTCTCGTTG 698
QY 217 AspGlnAlaPheAspLeuTyrCysArg 225
DB 699 GAAGAAGGGGTTGATCTGTATTGCAGG 725

```

# RESULT 11

CD834047 734 bp mRNA linear EST 10-JUL-2003  
 LOCUS BN45.040617F011018 BN45 Brassica napus cDNA clone BN45040617, mRNA  
 DEFINITION sequence.  
 ACCESSION CD834047  
 VERSION CD834047.1 GI:32515987  
 KEYWORDS EST.  
 SOURCE Brassica napus (rape)  
 ORGANISM Brassica napus  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
 REFERENCE 1 (bases 1 to 734)  
 AUTHORS Genoplatne, a major partnership french program in plant genomics  
 TITLE Unpublished (2003)  
 JOURNAL Contact: Genoplatne

Genoplatne  
 93, rue Henri Rochefort 91025 EVRY CEDEX France  
 Tel: 33 1 69 47 54 00  
 Fax: 33 1 69 47 54 10  
 This sequence has been generated in the framework of the french  
 plant genomics programme 'Genoplatne' (<http://www.genoplatne.com>  
 and <http://genoplatne-info.infobiogen.fr>).

## FEATURES

source  
 1..734  
 /organism="Brassica napus"  
 /mol\_type="mRNA"  
 /cultiivar="Jec Neuf"  
 /db\_xref="taxon:3708"  
 /clone="BN45040617"  
 /tissue\_type="seed"  
 /clone\_id="BN45"

## ORIGIN

### Alignment Scores:

Pred. No.:	1..03e-96	Length:	734
Score:	936.50	Matches:	188
Percent Similarity:	89.22%	Mismatches:	19
Best Local Similarity:	81.03%	Indels:	10
Query Match:	49.34%	Gaps:	5

US-10-019-931-3 (1-359) x CD834047 (1-734)

```

QY 1 MetAlaThrSerSerMetLySerIlePro---MetAlaIleProSerPheSerMetCys 19
DB 53 ATGGGACCTCAAGACATCAAGAGTCTTCATTAATGCGATCCCAAGTTCTCCATTGCG 112
QY 20 HisLysLeuGluLeuLeuLys---GluGlyLysThrArgAspValProLys-----Ala 36
DB 113 CACAAGCAGAGGCTCTCAAGAGAGAGCCGAGAC---CCGAAAGCCCAAGAA 169
QY 37 GluGluAspGluGlyLeuSerCysGluPheGlnGluMetLeuAspSerLeuProLysGlu 56
DB 170 GAAGAAGAAAGAGGGCTTAAGCTTAAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 229
QY 57 ArgGlyTPArgThrArgTyrLeuTyrLeuPheGlnGlyPheTyrCysGlnAlaLysGlu 76
DB 220 AGAGTTGGAGAACCTCGTCATCTTTACCTGTTCCAAAGCTTTTGGTGGCCAAAGCAGAG 289
QY 77 IleGlnAlaIleMetSerPheGlnLysHisPheGlnSerLeuGluAsnAspValValLeu 96
DB 290 ATTCATGCCATCATGTCTTCCAGAAACATTTCAAGACTCTCCCAAAAGATGTATTCTC 349
QY 97 AlaThrIleProLysSerGlyThrThrTrpLeuLysAlaLeuThrPheThrIleLeuAsn 116
DB 350 GGCACCATTAACCAAAATCCGGTACCAATGGTTAAAGCTTTAAACGTTTAAACCGCCTTAAC 409
QY 117 ArgHisArgPheAspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnPro 136
DB 410 CGACACCCGTTTGAATCCGTT-----TCGACCACTCTCTCTCACAACAACCTT 460
QY 137 HisAspLeuValProPhePheGluTyrLysLeuTyrAlaAsnGlyAspValProAspLeu 156
DB 461 CATGACCTCGTACCTTTTGGAGTACAAAGCTTTACGCCAAGAGAGATCTCGTATCTC 520
QY 157 SerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGlu 176
DB 521 TCCGGTCTAGCTAGTTCCAAGAACATTCGCAACACAGCTCTTCCGTTCTCTCAAGGGC 580
QY 177 ThrIleGluLysProGlyValLysValValTyrLeuCysArgAsnProPheAspThrPhe 196
DB 581 TCCATCGAAGAACCGGAGCGAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 640
QY 197 IleSerSerTPHisTyrThrAsnAsnIleLysSerGlySerValSerProValLeuLeu 216
DB 641 ATCTCTCTGTGCAATTCAGCAACAGCATTAAGTGGAGTCCGTTGAGTCCAGTCTCGTTG 699
QY 217 AspGlnAlaPheAspLeuTyrCysArg 228

```

Db 700 GAAGAAGGTT-GATCTATTTCAGGGAGTGATTA 734

RESULT 12

LOCUS AU238005 671 bp mRNA linear EST 01-APR-2002

DEFINITION AU238005 RAF16 Arabidopsis thaliana cDNA clone RAF16-75-P02 5', mRNA sequence.

ACCESSION AU238005

VERSION AU238005.1 GI:19877174

KEYWORDS EST.

SOURCE Arabidopsis thaliana (chale cress)

ORGANISM Arabidopsis thaliana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

AUTHORS 1 (bases 1 to 671)  
Seki M., Narusaka M., Ishida J., Kamiya A., Satou M., Nakajima M., Akiyama K., Enju A., Oono Y., Sakurai T., Carninci P., Kawai J., Itoh M., Ishii Y., Arakawa T., Shibata K., Shinagawa A., Muramatsu M., Hayashizaki Y. and Shibasaki K.

TITLE Large scale analysis of Arabidopsis full-length cDNA

JOURNAL Unpublished (2002)

COMMENT Contact: Motoaki Seki  
Plant Functional Genomics Research Group  
RIKEN Genomic Sciences Center  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-4359  
Fax: 81-298-36-9060  
Email: maseki@rc.riken.go.jp

FEATURES

source

1..671

location/Qualifiers

/organism="Arabidopsis thaliana"

/mol\_type="mRNA"

/db\_xref="taxon:3702"

/clone="RAF16-75-P02"

/lab\_host="DH10B"

/clone\_lib="RAF16"

/note="Site\_1: BamHI; Site\_2: SalI; dark-grown"

ORIGIN

Alignment Scores:

Pred. No.: 1.81e-94 Length: 671

Score: 916.50 Matches: 177

Percent Similarity: 90.14% Conservative: 15

Best Local Similarity: 83.10% Mismatches: 16

Query Match: 48.23% Indels: 5

DB: 1 Gaps: 2

US-10-019-931-3 (1-359) x AU238005 (1-671)

4 SerSerMetLysSerIleProMetAlaIleProSerPheSerMetCysHisLeuGlu 23

45 TCAAGCATTAAGAGAGATTCATGCGATCCCAATTTCTCATGTCACAGGCCGAG 104

24 LeuLeuLysGluGlyLysThrArgAspValProLysAlaGluGluAspGluLysSer 43

105 CTCCTTAAGCAAGCAAAAGC-----GAAGGCCAAGAAAGAGAGGCTTAAGC 152

44 CysGluPheGlnGluMetLeuAspSerLeuProLysGluArgGlyTTrpArgThrArgTyr 63

153 TACGAGTTCCAAAGATGTGSACTCTCTCTTAAGAGAGAGGACGGAGGAATGTTAC 212

64 LeuTyrLeuPheGlnGlyPheTrrPyrCysGlnAlaLysGluIleGlnAlaIleMetSerPhe 83

213 CTTTACTTATTCGAAGGTTTCGGTCCCAAGCTTAAGAGATTCAGGCTATCATCGTCTTC 272

Qy 84 GlnYerHisPheGlnSerLeuGluAsnAspValValLeuAlaTrpIleProLysSerGly 103

Db 273 CAAAACATTTTCAGTCCCTTCACAGCAGCTTTCCTCGCACACATACCTTAATCTGGC 332

Qy 104 ThrTrpTrpLeuLysAlaLeuThrPheThrIleLeuAsnArgHisArgPheAspProVal 123

Db 333 ACAACCTGGTTAAAGCTTTTAACCTTCACCATCTTCACCGGTACGTTGATCCGGTT 392

Qy 124 --AlaSerSerThrAsnHisPProLeuPheThrSerAsnProHisAspLeuValProPhe 142

Db 393 TCCATCATCAAGTTCGACACCCCTCTTCACATCCAAACCTTCACGACTCGTACCTTTC 452

Qy 143 PheGluTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAlaSerPro 162

Db 453 TTCAGATCAAGCTTTACGCCCAAGAAATTTCCGATCTCGGGCTAGCCAGTCCA 512

Qy 163 ArgThrPheAlaThrHisLeuProPheGlySerLeuLysGluTrpIleGluLysProGly 182

Db 513 AGAATTCGCAACCCACGATCCGTTGCGTGCCCTTAAGATTCGGTGAAGATCCAGT 572

Qy 183 ValLysValValTyrLeuCysArgAsnProPheAspThrPheIleSerSerTrpHisTyr 202

Db 573 GTGAAGGTTGTGATCCTGTGCGGACCCGTTTACACATTCATCTCCATGTCGATTAAC 632

Qy 203 ThrAsnAsnIleLysSerGluSerValSerProValLeu 215

Db 633 ATCAACCAATCATCTTCGATCGATGAGCCGACGCTTGG 671

RESULT 13

LOCUS CV130872 875 bp mRNA linear EST 03-SEP-2004

DEFINITION X9SP07a08 Populus stem seasonal library Populus deltoides cDNA, mRNA sequence.

ACCESSION CV130872

VERSION CV130872.1 GI:51874792

KEYWORDS EST.

SOURCE

ORGANISM

Populus deltoides

Populus deltoides

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids I; Malpighiales; Salicaceae; Salicaceae; Populus.

REFERENCE 1 (bases 1 to 875)  
Park S. and Han K.-H.

AUTHORS Gene expression profile during seasonal growth cycle in poplar tree

TITLE Unpublished (2003)

JOURNAL

COMMENT Contact: Kyung-Hwan Han  
Department of Forestry  
Michigan State University  
126 Natural Resources, East Lansing, MI 48824-1222, USA  
Tel: 517 353 4751  
Fax: 517 432 1143  
Email: hanky@msu.edu

FEATURES

source

1..875

location/Qualifiers

/organism="Populus deltoides"

/mol\_type="mRNA"

/strain="IL-129"

/db\_xref="taxon:3696"

/tissue\_type="stem"

/dev\_stage="1 Year old"

/clone\_lib="Populus stem seasonal library"

ORIGIN

Alignment Scores:

Pred. No.: 1.26e-92 Length: 875

Score: 902.00 Matches: 160

Percent Similarity: 75.62% Conservative: 54

Best Local Similarity: 56.54% Mismatches: 65

Query Match: 47.52% Indels: 4

DB: 7 Gaps: 1

US-10-019-931-3 (1-359) x CV130872 (1-875)

Qy	72	CyGlnAlaIyGluIleGlnAlaIleMetSerPheGlnIyPheGlnSerLeuGlu	91
Db	38	TCTCTGAAATGACAGATTCCTCTGATGATTCATTCACCACTCTTACACAGAA	97
Qy	92	AsnAspValValLeuAlaThrIleProIySerGlyThrThrIlePheValAlaLeuThr	111
Db	98	ACAGATTAATCTACTAGTCACATGCTTAATACAGACCAACATGTTGTAACCTTGACA	157
Qy	112	PheThrIleLeuAsnArgHisArgPheAspProValAlaSerSerThrAsnIleProLeu	131
Db	158	TTTTCATATATGAATGTCAGAAATATACACCTCTGTCAGC-----CCCTTG	205
Qy	132	PheThrSerAsnProHisAspLeuValProPhePheGluThrIleLeuValAlaLeuGly	151
Db	206	AACTCTGTCACACCTCATGATCTTGTACTTTCTTGAAGTTGGCTTACCCCAATATAC	265
Qy	152	AspValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrIleLeuProPhe	171
Db	266	CAACTCTCGACCTGCTACTCTTCCATCCCTAGATTTTGTACTCATCTGCATAT	325
Qy	172	GlySerLeuLeuGluThrIleGluIySProGlyValIyLeuValIyLeuCyAsnArgAn	191
Db	336	CCATCACTACCGGATTCATCAAGAACTCCGCTGTGCAATTGTTATCTTTCAGAGAA	385
Qy	192	ProPheAspThrPheIleSerSerTrpHisArgThrAsnAsnIleIySerGlySerVal	211
Db	366	CCTTTTGACAACTTATCTCTTGTGGCATTTCCGCTCCAAAGCAACATGAAAGTCT	445
Qy	212	SerProValLeuLeuAspGlnAlaPheAspLeuThrCyAsnArgIyValIleGlyPheGly	231
Db	446	GCGCATTAATCTTTGGAGGATTTGTCATGATGTTTGCATGCAATGCACTGAGCATTCGCT	505
Qy	232	ProPheTrpGluHisMetLeuGlyThrTrpArgIySerLeuIyAsnArgProGluIySVal	251
Db	506	CCCTTTTGACAGATTAATGAGGATTTGAGAGCTTGAAGACCAAGAGAGGTT	565
Qy	252	PhePheLeuArgTrpGluAspLeuIySAspAspIleGluThrAsnLeuIySArgLeuAla	271
Db	566	CGTTTCTCAGCATGAGACATGAAGAAAGACATTAATCTCAAGTAAAGGCTAGCT	625
Qy	272	ThrPheLeuGluLeuProPheThrIleGluIyGluIySArgIyValIyAlaIle	291
Db	626	GAGTTCCTGGGCTGCTCTTTTCTTGGAAAGAGGCGATGGGTTGTGCAAGAAAT	685
Qy	292	AlaGluLeuCySerPheGluAsnLeuIySLeuGluValAlaAsnIySerAsnIyS	311
Db	686	TCAAGTCTGTAGCTTCAGCAATTTGAAGACAAAGATCAACAGACTGCGAAGTCT	745
Qy	312	IleIyAsnPheGluAsnArgPheLeuPheArgIySArgIyValIySerAspTrpValAsn	331
Db	746	ATCCCACTATGAGAAACAAAGACTCTTTAAGAAAGGTGAAGTGGGATTTGGGTCAAT	805
Qy	332	ThrIleSerProSerGlnValGluArgLeuSerAlaLeuValAspAspIySLeuGly	351
Db	806	TACCTTAATCTCCGATGATGATGATGATTTGAACAAATATCAGCAACAAAGCTGCTGCT	865
Qy	352	SerGlyLeu	354
Db	866	TCTGTTTG	874
RESULT 14			
LOCUS	CX043920	817 bp	mRNA linear EST 09-DEC-2004
DEFINITION	UCRCS07_14C09 g Parent Washington Navel Orange Thrip-Challenged		
	Flavado cDNA library UCRCS07 Citrus sinensis cDNA clone		
	UCRCS07-14C09-F18-1-7.g, mRNA sequence.		
ACCESSION	CX043920		
VERSION	CX043920.1	GI:56528211	
KEYWORDS	EST.		
SOURCE	Citrus sinensis		
ORGANISM	Citrus sinensis		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		

REFERENCE	1 (bases 1 to 817)	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; euroside II; Sapindales; Rutaceae; Citrus.
AUTHORS	Morse, J., Roose, M. L., Federici, C. F., Fenton, R. D., Watkins, P., Ikeda, J., Collin, M., Kacari, Y., Landry, B., Hubert, N., Laforet, M., Landry, J., and Ligonde, A.	
TITLE	Development of EST Resources and New Genetic Markers for California Citrus - Parent Washington Navel Orange Thrip-Challenged Flavado cDNA library UCRCS07	
JOURNAL	Unpublished (2004)	
COMMENT	Contact: Timothy J. Close Department of Botany & Plant Sciences University of California Riverside, CA 92521-0124, USA Tel: 909-787-3318 Fax: 909-787-4437 Email: timothy.close@ucr.edu Seq primer: T3.	
FEATURES	source	
	Location/Qualifiers	
	1..817	
	/organism="Citrus sinensis"	
	/mol_type="mRNA"	
	/cultivar="Parent Washington Navel"	
	/db_xref="taxon:2711"	
	/clone="UCRCS07-14C09-F18-1-7.g"	
	/tissue_type="Flavado"	
	/dev_stage="11 year old trees"	
	/lab_host="E. coli TG121"	
	/clone_lib="Parent Washington Navel Orange Thrip-Challenged Flavado cDNA library UCRCS07"	
	/note="vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI. Trees were grown in the field at University of California, Riverside using standard horticultural practices. Young fruits were placed in a cage with thrips (Scirtothrips citri). Infestations were conducted by Watkins (Morse lab) and Flavado collected by Federici (Roose lab) from May-June 2003. The thrips were collected from naturally infested Rhus plants by sucking into a tube. A flexible hose was attached to a tube that extended into a covered vial, and another tube stuck out of the vial at a right angle. The bent tube was held above the thrips, sucking on the flexible tube created a vacuum, pulling the thrips into the vial. Thrips were knocked off the Rhus plant onto a manila folder, then only second instars were captured. This was done repeatedly until enough were obtained. Approximately 7-10 thrips were caged on each fruit within a plastic vial made of a 8 cm long by 5 cm diameter tube that had a very fine mesh organically fabric glued to the bottom. The plastic cap was slit from the edge to the center so it could be slipped over the stem of the fruit. It was put in place, the thrips were knocked into the vial and it was fastened onto the cap then all gaps were closed with masking tape. The thrips naturally move up to the fruit. The cages were left in place for two days, then removed. The fruit were checked to be sure the thrips had stayed on, and then brought to the lab to cut off the flavado using a razor blade. Only the flavado from the stem 1/3 to 1/2 of the fruit was used. For controls an equal number of comparable sized fruit were caged without thrips, and the peel collected from them in the same manner. Tissues were frozen in liquid nitrogen, then stored at -80C until further processing. Fenton (Close lab) purified RNA by the phenol method described in J. Japanese Soc. Hort. Sci. 1996. 64 (4): 809-814, purified poly(A) mRNA using an Oligotex mRNA Kit (Qiagen), produced a primary cDNA library using a lambda ZAP XR cDNA Synthesis Kit (Stratagene), then mass-excised 0.77 million pfu from the primary library to produce a phagemid population. The library was made from equal portions of RNA from each of the two treatments. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally	

using an B13730 at DNA landmarks (Landry, Hubert, Laforest, Landry, Ligonde). Chromatogram files were downloaded by FTP by Cloose, then processed by Manamaker (Cloose lab) using the HARVEST pipeline (<http://harvest.ucr.edu/>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Cloose, Kooser, Federick, Manamaker, Lyon, Je, Jang, Collin, Kacar, Ikeda, Quintillo). Sequences that survived all removal steps were submitted to GenBank."

**ORIGIN**

### Alignment Scores:

Pred. No.:	5.29e-91	Length:	81
Score:	887.50	Matches:	16
Percent Similarity:	77.69%	Conservative:	42
Best Local Similarity:	61.54%	Mismatches:	53
Query Match:	46.76%	Indels:	5
DB:	8	Gaps:	3

US-10-019-931-3 (1-359) x CX043920 (1-817)

QY	30	ThzrAgapValProLysAlaGluGlnAspGluGlyLeuSerCysGluPheGlnGluMet	45
Db	34	ACAAACAACCCAAATAGACAGATGAAGAAGAAACCACTTAAGCAGTGAAGCCAAAGATGTG	93
QY	50	LeuAapSerLeuProLysGluArgGlyTrpArgThrArgTyrLeuTyrLeuPheGlnGly	69
Db	94	CTTCCTTCACTTCCCTTAAGAAAAGAGTTGGCCAGCGCTTTCTTTATAGTTCCAGGG	153
QY	70	PheTrpCysGlnAlaLysGluIleGlnAlaIleMetSerPheGlnLysHisPheGlnSer	89
Db	154	TATTGGTGCAGGCAAGAGAGATTCAAGCATATAGTGCCTTCCAAAAGCACTTCAAGCA	213
QY	90	LeuGluAenAapValValLeuAlaThrIleProLysSerGlyThrThrTrpLeuValAla	109
Db	214	AAAGCACTGATATCACTTTTATGCCAGATACCAAAATACAGGCCACTTGGATGAAGCA	273
QY	110	LeuThrPheThrIleLeuAenAargHisArgCysPheAspProValAlaSer-----Ser	126
Db	274	TTGTCTTTGGCAATATATCAATGCGCAAGAAATTTC--CCCATATTTAGTATCATCATGCT	330
QY	127	ThrAenHisProLeuPheThrSerAenProHisAspLeuValProPhePheGluTyrLys	146
Db	331	CATCATCAACCCCTGGCTTACTTCAATTCCTCAATGATCTTGCTCTTTCTTGAATACAAG	390
QY	147	LeuTyrAlaAenGlyAspValProAapLeuSerGlyLeuAla--SerProArgThrPhe	165
Db	391	CTCTATGCAAAATAAACCAATTCCTGTGACTTTCCTCAATTCGTGATGAGCCCTTAAGTTTT	450
QY	166	AlaThrHisLeuProPheGlySerLeuLysGluThrIleGluLysProGlyValLysVal	185
Db	451	GCCACCCCAATTCACATTTGCTCTCTTAAATTTGCTTCTTCAATGAACAACATTAAGATT	510
QY	186	ValTyrLeuCysArgAenProPheAapThrPheIleSerSerTrpHisTyrThrAenAsn	205
Db	511	GTTTATATTGGAGAAACCCTTTTGAATACTTTCATCTCTTCATGCGCATTTTCTTAACAA	570
QY	206	IleLysSerGlnSerValSerProValLeuLeuAspGlnAlaPheAapLeuTyrCysArg	225
Db	571	TTGAGGTCTCAAGGATTAACCTGAAATTTCACTGAGGAAGCAATTCAAAATGTACTGGAT	630
QY	226	GlyValIleGlyPheGlyProPheTrpGluHisMetLeuGlyTyrTrpArgGlnSerLeu	245
Db	631	GGTGTCATTGGGTGGCTCATTTTGGAGCAATATTTAGGGATTGGAAATGAGAGCTTG	690
QY	246	LysArgProGluLysValPhePheLeuArgTyrGluAspLeuLysAapIleGluThr	265
Db	651	AAGAGACCAACAATGTCTGTGTTTGAAGTATGACGATATGAAACAAGCATTTGTCTCA	750
QY	266	AsnLeuLysArgLeuAlaThrPheLeuGlnLeuProPheThrGluGlnGluGluArgLys	285

Db 751 AATTGAGAAATGGCAAGCTTTTGGGGTCCCTTTTCCCTGAGGAGGAGACAAG 810

## RESULT 15

AV827990

## DEFINITION

**ACCESSION**

**VERSION**

## KEYWORDS

**SOURCE**

## ORGANIS

## DEPENDENCE

## REFERENCES

## AUTHORS

**TITLE**

## FEATURES

**Source**

ORIGIN

**Alignment Scores:**

Pred. No.:	3.24e-89	Length:	574
Score:	870.00	Matches:	169
Percent Similarity:	98.26%	Conservative:	0
Best Local Similarity:	98.26%	Mismatches:	1
Query Match:	45.84%	Indels:	0
DB:	1	Gaps:	0

US-10-019-931-3 (1-359) X AV827990 (1-574)

Qy	1	MetAlaIthSerSermetLysSerIleProMetAlaIleProSerPheSerMetCysHis	20
Db	63	ATGGTACTCAAGCATGAGAGCATTCACATGGGAGATCCCAATTCTCCATGTGTAC	122
Qy	21	LysLeuGluLeuLeuLysGluGlyIysThrArgAspValProLysAlaGluGluAspGlu	40
Db	123	AAGCTGACACTCTTAAAGAAAGGCAAAATCGCGACGTCCGAAAGCCGAAGAAGTAA	182
Qy	41	GlyLeuSerCysGluPheGlnGluMetLeuAspSerLeuProLysGluArgGluTyrArg	60
Db	183	GGGCTAAGTCGAGAGTCCAAAGAGATGTGAAATCTCTCTTAAGAGAGAGAGATGAGA	242

```
Qy 61 ThrArgTyrLeuTyrLeuPheGlnGlyPheTrpCysGlnAlaLysGlnIleGlnAlaIle 80
Db 243 ACTCGTTACCTTACCTAATCCAAAGGGTTTGGTGCCAAAGCCAAAGAGATTCAGCCATC 302
Qy 81 MetSerPheGlnLysHisPheGlnSerLeuGluAmaAspValValLeuAlaThrIlePro 100
Db 303 ATGCTTTCCAAAACATTTCCAAATCCCTCGAAAAGAGTGCTTCGCCACCATACCT 362
Qy 101 LysSerGlyThrThrTyrPleuLysAlaLeuThrPheThrIleLeuAmaArgHisArgPhe 120
Db 363 AAATCCGTAACAACCTGGCTAAAGCTTTAATTACCATCTTAACCGTCACCGGTTT 422
Qy 121 AspProValAlaSerSerThrAmaHisProLeuPheThrSerAsnProHisAspLeuVal 140
Db 423 GATCGGTTGGCTCGAGTACCAACCAACCCTTTCACTTCCAAACCTCATGACCTTGT 482
Qy 141 ProphePheGluTyrLysLeuTyrAlaAsn-GlyAspValProAspLeuSerGly-Leu 160
Db 483 CTTTCTTCGAGTACAGCTTACGCCCAACCGAGATGTTCCCGATCTCTCNGGTTCTAG 542
Qy 160 LysSerProArgThrPheAlaThrHisLeuPro 170
Db 543 CCAGTNCAGAAAGTTCCCAACCCCACTTACCG 574
```

Search completed: December 23, 2005, 21:42:32  
Job time : 3732 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: December 23, 2005, 18:54:37 ; Search time 214 Seconds  
(without alignments)  
2981.984 Million cell updates/sec

Title: US-10-019-931-3  
Perfect score: 1898  
Sequence: 1 MATSMKSIIPMAIPFSWCH.....RLSALVDDKIGSGHTRFLS 359

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPFO\_spool\_p/US10019331/runat\_16122005\_165219\_18469/app\_query.fasta\_1.519  
-DB=Issued\_Patents\_NA -QFWT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human4.0.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10019331@cgn2\_1.1.290@runat\_16122005\_165219\_18469 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC  
-DEV TIMEOUT=130 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELXT=7

Database :

Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/1/ina/1\_COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5\_COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/H\_COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/PCITUS\_COMB.seq: \*  
7: /cgn2\_6/ptodata/1/ina/PP\_COMB.seq: \*  
8: /cgn2\_6/ptodata/1/ina/RE\_COMB.seq: \*  
9: /cgn2\_6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	550.5	29.0	1192	3	US-09-854-122-15 Sequence 15, Appl
2	325.5	17.1	922	3	US-09-609-816-2 Sequence 2, Appl
3	325.5	17.1	922	3	US-10-199-334-2 Sequence 2, Appl
4	325.5	17.1	922	3	US-10-199-329-2 Sequence 2, Appl
5	300	15.8	1004	3	US-09-609-816-3 Sequence 3, Appl
6	300	15.8	1004	3	US-10-199-334-3 Sequence 3, Appl
7	300	15.8	1004	3	US-10-199-329-3 Sequence 3, Appl
8	296.5	15.6	942	3	US-09-609-816-1 Sequence 1, Appl
9	296.5	15.6	942	3	US-10-199-334-1 Sequence 1, Appl

10	296.5	15.6	942	3	US-10-199-329-1	Sequence 1, Appl
11	294	15.5	1396	3	US-09-328-1174A-2	Sequence 2, Appl
12	286.5	15.1	1045	3	US-09-949-016-4457	Sequence 4457, Ap
13	286.5	15.1	1063	2	US-08-325-562-1	Sequence 1, Appl
14	286.5	15.1	1063	2	US-08-437-795-1	Sequence 1, Appl
15	278.5	14.7	993	3	US-09-717-321A-45	Sequence 45, Appl
16	275	14.5	912	3	US-09-795-926-1	Sequence 1, Appl
17	275	14.5	912	3	US-10-364-774-1	Sequence 1, Appl
18	266.5	14.0	798	3	US-09-795-926-5	Sequence 5, Appl
19	266.5	14.0	798	3	US-10-364-774-5	Sequence 5, Appl
20	263	13.9	2431	3	US-09-949-016-147	Sequence 147, App
21	244.5	12.9	1447	3	US-09-949-016-5340	Sequence 5340, Ap
22	244.5	12.9	1477	3	US-09-949-016-655	Sequence 655, App
23	241	12.7	1130	3	US-09-786-240-22	Sequence 22, Appl
24	239.5	12.6	2153	3	US-09-795-926-19	Sequence 19, Appl
25	239.5	12.6	2153	3	US-10-364-774-19	Sequence 19, Appl
26	227.5	12.0	645	3	US-09-795-926-15	Sequence 15, Appl
27	227.5	12.0	645	3	US-10-364-774-15	Sequence 15, Appl
28	209	11.0	1196	3	US-09-949-016-2593	Sequence 2593, Ap
29	197	10.4	1564	2	US-08-852-481-1	Sequence 1, Appl
30	144.5	7.6	561	3	US-09-795-926-11	Sequence 11, Appl
31	144.5	7.6	561	3	US-10-364-774-11	Sequence 11, Appl
32	136	7.2	447	3	US-09-795-926-7	Sequence 7, Appl
33	136	7.2	447	3	US-10-364-774-7	Sequence 7, Appl
34	113.5	6.0	289	3	US-09-490-609B-340	Sequence 340, App
35	109.5	5.8	4411529	3	US-09-103-840A-1	Sequence 1, Appl
36	108.5	5.7	304	3	US-09-172-711-16	Sequence 16, Appl
37	108.5	5.7	304	3	US-10-126-279-24	Sequence 24, Appl
38	108.5	5.7	978	3	US-10-286-606-24	Sequence 24, Appl
39	106.5	5.6	4403765	3	US-09-103-840A-2	Sequence 2, Appl
40	103.5	5.5	924	3	US-09-328-352-2420	Sequence 2420, Ap
41	103.5	5.5	1485	2	US-08-468-036-8	Sequence 8, Appl
42	103.5	5.5	1485	2	US-08-376-843-8	Sequence 8, Appl
43	103.5	5.5	3098	2	US-08-447-500-1	Sequence 1, Appl
44	103.5	5.5	3098	2	US-08-454-097-1	Sequence 1, Appl
45	103.5	5.5	3098	2	US-08-447-408-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-09-854-122-15  
; Sequence 15, Application US/09854122  
; Patent No. 6641718  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTE, RANDALL S.  
; APPLICANT: SMITH, ROBERT  
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA  
; FILE REFERENCE: PHA-007.01  
; CURRENT APPLICATION NUMBER: US/09/854,122  
; CURRENT FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/202,529  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 1192  
; TYPE: DNA  
; ORGANISM: Zostera marina  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1..33, 37..1041, 1045..1059, 1063..1077,  
; LOCATION: 1081..1086, 1090..1119, 1123..1191)  
US-09-854-122-15

#### Alignment Scores:

Pred. No.: 9,38e-57  
Score: 550.50  
Percent Similarity: 57.91%  
Best Local Similarity: 36.71%  
Query Match: 29.00%  
DB: 3  
Length: 1192  
Matches: 116  
Conservative: 67  
Mismatches: 122  
Indels: 11  
Gaps: 8

[illegible]

```

: APPLICANT: Kodira, Chinappa
: APPLICANT: Beasley, Ellen
: APPLICANT: DiFrancesco, Valentina
: TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEIN, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USBS THEREOF
: FILE REFERENCE: CLO00699ECT
: CURRENT APPLICATION NUMBER: US/09/609,816
: PRIOR APPLICATION NUMBER: 60/192,408
: PRIOR FILING DATE: 2000-03-27
: PRIOR APPLICATION NUMBER: 60/212,725
: PRIOR FILING DATE: 2000-06-20
: PRIOR APPLICATION NUMBER: 09/609,816
: PRIOR FILING DATE: 2000-07-03
: NUMBER OF SEQ ID NOS: 23
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 922
: TYPE: DNA
: ORGANISM: HUMAN
US-09-609-816-2

Alignment Scores:
Pred. No.: 1,4e-29 Length: 922
Score: 325.50 Matches: 86
Percent Similarity: 46.98% Conservative: 52
Best Local Similarity: 29.25% Mismatches: 101
Query Match: 17.15% Indels: 55
DB: 3 Gaps: 8

US-10-019-931-3 (1-359) x US-09-609-816-2 (1-922)

QY 86 HispHeGInSerLeuGluAaMhspValValLeuAlaThrIleProLySerserGlyThrThr 105
Db 128 AATTTCCAGGCCAAGCCCGATGATCTTATTTCTGGCAACTTACCCAAGTCAGGTACACAA 187
QY 106 TrpLeuYsAlaLeuThrPheThrIleLeuAaen----- 116
Db 168 TGGATGATGAAATTTTGAACATGATCTTAATTAATGATGCGTGAATGGAGAAATGCCAAAGA 247
QY 117 -----ArgHisarGpheaSpProValAlaSerSerThrAaHsIspProLeu 131
Db 248 GCCCAGACTCTAGATAGACACCCCTTC-----CTT 277
QY 132 PheThrSerAsnProHisaspLeuValProPhepHeGlyTrLyLeuYrAlaenGly 151
Db 278 GAATCGAATATTTCCCATTAAGAAACCAAGATTGGAGTTGCTTCTT----- 325
QY 152 AspValProaspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPhe 171
Db 326 -----GAAATGTCCTCACCAACAACCTGATTAACACATCTCCCTTCA 367
QY 172 GlySerLeuYsGlyThrIleGlyLysProGlyValYsValValYrXLeuCYsArGaaen 191
Db 368 CATCTGATTTCCACCATCTATCTGAAAGAAACCTGAAATATGTCTATGTGGCCAGAAAT 427
QY 192 PropheaSpThrPheIleSerSerTrpHisYrThrAaenAlleYsSerGlySerVal 211
Db 428 CCCAAGATGGCTGGTGCTCTACTACCACTTT---CACAGATGGCTTCTTTATGCGCT 484
QY 212 SerProValLeuLeuaspGlnAlaPheaspLeuYrCYsArgGlyValIleGlyPheGly 231
Db 485 GATCTCTAGAACTTGAAGAAATTTATATGAAATTCATCTCCGAGAAAGTTGTGGCGGG 544
QY 232 ProPheTrpGlnHisMetLeuGlyYrTrpArgGluSerLeuYsArgProGlyYsVal 251
Db 545 TCCTGCTTTGACCACTGTAAGAGATGCTG-----GCTGCCAAAGACATGACCGGATC 598
QY 252 PhepHeLeuArgTrGlyLaspLeuYsAspAspIleGlyThrAaenLeuYsArgLeuAla 271
Db 599 CTTACTCTTCTTACAGAGATATTTAAACCAAGCCAAACGGGAAATTTGAGAGATATCTG 658

```



```

QY 272 ThrPheLeuGluLeuProPheThrgluGluGluArgLyGlyValValLysAlaIle 291
DB 659 AAGTTCCTGGAAAAAGACATATCAGAGAA-----ATTCTGATATAAATC 703
QY 232 AlaGluLeuCySerPheGluAsnLeuLys-----Leu 303
DB 704 ATCTATCAACACTCCTTGTATGTATGAAGCAAAACCAATGACCAACTATCACTTTG 763
QY 304 GluValAsnLysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeuPheArgLys 323
DB 764 CCCACACGATTATGACACCTCCATCTCCCTTT-----ATGAGGAAA 808
QY 324 GlyGluValSerAspTrpValAsnTrpLeuSerProSerGlnValGluArgLeuSerAla 343
DB 809 GGGATGCTCGACGACGTGAAGAACTATTTTACTGTGGCCCAAAATGAAATTTGACAAAG 868
QY 344 LeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArg 357
DB 869 GACTACCAAGAAAGATGGCAGAGAACCCCTAACCTTCCGC 910

RESULT 3
US-10-199-334-2
; Sequence 2, Application US/10199334
; Patent No. 6905855
; GENERAL INFORMATION:
; APPLICANT: WOODAGE, Trevor et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CLO00669DIV-3
; CURRENT APPLICATION NUMBER: US/10/199,334
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 922
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-199-334-2

Alignment Scores:
Pred. No.: 1,4e-29 Length: 922
Score: 325.50 Matches: 86
Percent Similarity: 46.94% Conservatave: 52
Best Local Similarity: 29.25% Mismatches: 101
Query Match: 17.15% Indels: 55
DB: 3 Gaps: 8

US-10-019-931-3 (1-359) x US-10-199-334-2 (1-922)
QY 86 HisPheGlnSerLeuGluAsnAspValLysAlaThrIleProLysSerGlyThrThr 105
DB 128 AATTTCCAAAGCAACCGCTGATGATCTTATTCGGCAACTTACCAAGTCAGSTACACA 187
QY 106 TrpLeuLysAlaLeuThrPheThrIleLeuAsn----- 116
DB 188 TGGATGCAATGAATTTTACACATGATTTCTAAATGATGTGATGTGAGAAATGCAAAAGA 247
QY 117 -----ArghIsArgPheAspProValAlaSerSerThrAsnHisProLeu 131
DB 248 GCCCAGACTCTGATATGACACCGCTTTC-----CTT 277
QY 132 PheThrSerAsnProHisAspLeuValProPhePheGluTrpLysLeuTrpLysAlaSerGly 151
DB 278 GAACGGAATATTTCCCATATAAGAAAAACCAAGATTGGATTTGCTTCTT----- 325

```

```

QY 152 AspValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPhe 171
DB 326 -----GAAATGCTCCTACCAACATCATGATATAAACAACATCTCCCTTCA 367
QY 172 GlySerLeuLysGluThrIleGluLysProGlyValLysValValTrpLeuCyAsnArgAsn 191
DB 368 CATCTGATTTCCACCATCTATCTGAGAAAGAAACCAAGATTTGTCTATGTGGCCAGAAAT 427
QY 192 ProPheAspThrPheIleSerSerTrpHisTrpThrAsnAsnIleLysSerGluSerVal 211
DB 428 CCCAAGGATTCCTGTGCTCTACTACCACTTT---CACAGATGCTTCTTTATGCTCT 484
QY 212 SerProValIleLeuAspGlnAlaPheAspLeuTrpCyAsnGlyValIleGlyPheGly 231
DB 485 GATCTCAGAACTTATGAGAAATTTATAGAAATTCATGTCGGGAAAAGTTGTGGCGGG 544
QY 232 ProPheTrpGluHisMetLeuGlyTrpTrpArgLysSerLeuLysArgProGluLysVal 251
DB 545 TCCTGCTTGACATGTCGAAAGATGTG-----GCTGCAAAAGACATGACCGGATC 598
QY 252 PhePheLeuArgTrpGluAspLeuLysAspAspIleGluThrAsnLeuLysArgLeuAla 271
DB 599 CTCTACCTCTTCTACGAGATATTTAAAAAGCCCAAGCGGGAATTGAGAAAGATGACTG 658
QY 272 ThrPheLeuGluLeuProPheThrgluGluGluArgLyGlyValValLysAlaIle 291
DB 659 AAGTTCCTGGAAAAAGACATATCAGAGAA-----ATTCTGATATAAATC 703
QY 232 AlaGluLeuCySerPheGluAsnLeuLys-----Leu 303
DB 704 ATCTATCAACACTCCTTGTATGTATGAAGCAAAACCAATGACCAACTATCACTTTG 763
QY 304 GluValAsnLysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeuPheArgLys 323
DB 764 CCCACACGATTATGACACCTCCATCTCCCTTT-----ATGAGGAAA 808
QY 324 GlyGluValSerAspTrpValAsnTrpLeuSerProSerGlnValGluArgLeuSerAla 343
DB 809 GGGATGCTCGACGACGTGAAGAACTATTTTACTGTGGCCCAAAATGAAATTTGACAAAG 868
QY 344 LeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArg 357
DB 869 GACTACCAAGAAAGATGGCAGAGAACCCCTAACCTTCCGC 910

RESULT 4
US-10-199-329-2
; Sequence 2, Application US/10199329
; Patent No. 6953681
; GENERAL INFORMATION:
; APPLICANT: WOODAGE, Trevor et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CLO00669DIV-2
; CURRENT APPLICATION NUMBER: US/10/199,329
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 922
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-199-329-2

```

## Alignment Scores:

Pred. No.:	1,4e-29	Length:	922
Score:	325.50	Matches:	86
Percent Similarity:	46.94%	Conservative:	52
Best Local Similarity:	29.25%	Mismatches:	101
Query Match:	17.15%	Indels:	55
DB:	3	Gaps:	8

US-10-019-931-3 (1-359) x US-10-199-329-2 (1-922)

```

QY      86  H1sPheGlnSerLeuGlnAspValValLeuAlaThrLeProLysSerGlyThrThr 105
      128  AATTTCAGCCCAAGCGCTGATGATCTTAATCTCGCACTTACCAAGTCAGGTACACA 187
QY      106  TTPLeuValAlaLeuThrPheThrIleLeuAsn----- 116
      188  TGGATGTCATGAATTTTAAAGATGATTTAAAGATGGTGAATGAGAAATGCAAAAGA 247
QY      117  -----ArgHisArgPheAspProValAlaSerSerThrAsnHisProLeu 131
      248  GCCCAGACTCTAGATGACACGCTTC-----CTT 277
QY      132  PheThrSerAsnProHisAspLeuValProPheGlnGlyTrLysLeuTyAlaAsnGly 151
      278  GAACCTGAATTTCCCATTAAGAAAAACAGATTTGAGATTCGTCCTT----- 325
QY      152  AspValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPhe 171
      326  -----GMAATGTCCTCACCAACCTGATAAAACACATCTCCCTTCA 367
QY      172  GlySerLeuGlnGlyThrIleGlyLysProGlyValIleValValTyLeuCysArgAsn 191
      368  CATCTGATTCACCATCTATCTGGAAGAAAGAAACATCGAAGTTGCTCATGCGCCAGAAAT 427
QY      192  ProPheAspThrPheIleSerSerTrpHisGlyTrpThrAsnAsnIleLysSerGlnSerVal 211
      428  CCCAAGATGCTGCTGCTCTACTACCACTT---CACAGATGGCTTCCTTATGCTT 484
QY      212  SerProValLeuLeuAspGlnAlaPheAspLeuTyCysArgGlyValIleGlyPheGly 231
      485  GATCTCCAGAACTTAAAGGAATTTTAAAGAAATTCATGTCGGAAGAAAGTTGGCGGG 544
QY      232  ProPheTrpGlnHisMetLeuGlyTyTrpArgGlnSerLeuLysArgProGlnLysVal 251
      545  TCTGTGTTGACCATGTCGAAGAGATGCTG-----GCTGCAAAAGCATGACCGGATC 598
QY      252  PhePheLeuArgTyTrpGlnAspLeuLysAspAspIleGluThrAsnLeuLysArgLeuAla 271
      599  CTCTACCTCTTCTACGAGATATTAAAAAGACCAAAAGCGGAATTTGAGAAATTACTG 658
QY      272  ThrPheLeuGlnLeuProPheThrGlnGlnGlnGlnArgGlyValValIle 291
      659  AAGTTCCTGAAAAAGACATATCAGAGAA-----ATTCTGAATTAATATC 703
QY      292  AlaGlnLeuCysSerPheGlnAsnLeuLysLys-----Leu 303
      704  ATCTATACACACCTCTTGTATGTAATGACAAACCAATGACCAACTATACCACTTGG 763
QY      304  GluValAsnLysSerAsnLysSerIleLysAsnPheGlnAsnArgPheLeuPheArgLys 323
      764  CCCACAGCGCATTTATGACCACTCTCCCTCTTT-----ATGAGGAAA 808
QY      324  GlyGlnValSerAspTrpValAsnTyLysSerProSerGlnValGlnArgLeuSerAla 343
      809  GGGATGCTGGAACCTGGAAGAACTATTTAATCTGTGCGCCCAAAATGAAATTTGACAG 868
QY      344  LeuValAspAspLysLeuGlySerGlySerGlyLeuThrPheArg 357
      869  GACTACCAAGAAAGATGCGAGAGACCACTTAACCTTCGCG 910

```

RESULT 5  
US-09-609-816-3  
Sequence 3, Application US/09609816

```

/ Patent No. 643684
/ GENERAL INFORMATION:
/ APPLICANT: Woodage, Trevor
/ APPLICANT: Wei, Minh Hui
/ APPLICANT: Kodira, Chinnappa
/ APPLICANT: Bealeley, Ellen
/ APPLICANT: Difrancesco, Valentina
/ TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
/ TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
/ TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
/ FILE REFERENCE: CL00069PCT
/ CURRENT APPLICATION NUMBER: US/09/609,816
/ PRIOR FILING DATE: 2000-07-03
/ PRIOR APPLICATION NUMBER: 60/192,408
/ PRIOR FILING DATE: 2000-03-27
/ PRIOR APPLICATION NUMBER: 60/212,725
/ PRIOR FILING DATE: 2000-06-20
/ PRIOR APPLICATION NUMBER: 09/609,816
/ NUMBER OF SEQ ID NOS: 23
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 3
/ LENGTH: 1004
/ TYPE: DNA
/ ORGANISM: HUMAN
US-09-609-816-3

```

## Alignment Scores:

Pred. No.:	2.01e-26	Length:	1004
Score:	300.00	Matches:	95
Percent Similarity:	42.82%	Conservative:	63
Best Local Similarity:	25.75%	Mismatches:	121
Query Match:	15.81%	Indels:	90
DB:	3	Gaps:	14

US-10-019-931-3 (1-359) x US-09-609-816-3 (1-1004)

```

QY      16  PheSerMetCysHisIleLysLeuGlnLeuLysGlnGlyLysThrArgAspValProLys 35
      38  TTGGCCCTTATGGCGAAGATTGAG-----AAAAACGCTCCACAG 76
QY      36  AlaGlnGlnAspGlnGlyLeuSerCysGlnPheGlnGlnMet---LeuAspSerLeuPro 54
      77  ATGCAAAAAAAGCCAGAACTG-----TTTAACATCATGAACTGATGAGATCGCT 127
QY      55  LysGlnArgGlyTrpArgThrArgTyTrpLeuTyLysPheGlnGlyPheTrp----- 71
      128  ACG-----TTGATATTATTCAAAAGAAATGATGGGAAAAAGATC 163
QY      72  CysGlnAlaLysGlnIleGlnAlaIleMetSerPheGlnLysHisAspGlnSerLeuGln 91
      164  TGT-----AATTCACAGCCAGCGCT 184
QY      92  AsnAspValValLeuAlaThrIleProLysSerGlyThrThrTrpLeuValAlaLeuThr 111
      185  GATGATATTCTTCTGCACTTACCCAAAGTCAGGTACAAACATGATGATGAATTTTA 244
QY      112  PheThrIleLeuAsn-----Arg 117
      245  GACATGATTTCTAAATATAGTGTGATGTCGAGAAATGCAAAAGAGCCAGACTTAATAGA 304
QY      118  HisArgPheAspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHis 137
      305  CACGCTTTC-----CTTGAACGTAAATTTCCCAT 334
QY      138  AspLeuValProPhePheGlnTyTrLysLeuTyAlaAsnGlyAspValProAspLeuSer 157
      335  AAAGAAAAACAGATTTGAGGTTCGTTCTT----- 364
QY      158  GlyLeuAlaSerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGlnThr 177
      365  GAAATGCTCTACCAACATGATTAACACATCTCCCTTACATGTGATTCACCATCT 424

```

Oy	178	llegluylsPrcGlyVallyleVallyrleuCyArAsnpProheAsprrPheille	197
Db	425	ATCTGGAAAGAAACTGCAGATGCTCTATGTGGCCAGAAATCCCAAGATTGCTGGTG	484
Oy	198	SerserrPhtslyrYrThraenAnileuysSergJuservalserProvalleuAsp	217
Db	485	TCCTACTACCACTTT---CACAGATAGGCTTCCTTATGCTCGATCTCTCAGAACTTAGAG	541
Oy	218	GlnAlapheAspleuYrCyArsglyValilleglyPheglyProphetrgJuhismet	237
Db	542	GAATTTTGTGAAGAAATTCATGTCCGGAAAGTGTGGCGGGCTCGGTTGACCATGTG	601
Oy	228	LeuglyYrTrpArgJusleuYsArGPrGJuleValPhePheleuArgYrGlu	257
Db	602	AAAGAGTGGTGG---GCTGCAGAAAGACAGCCAGCGGATCCTTACTTCTACGAG	655
Oy	258	AspleuYsAspAspillegluThraenleuYsArgleuAlatrPheleugJuleuPro	277
Db	656	GATATTAAAAAATCCAAACATGAGATCCCAAGGTGTGGAAATCTTGGAGAAACTG	715
Oy	278	PhehrJugluGluGluArglysglyVallyVallysalilleAlaGluLeuCysserPhe	297
Db	716	TTGTGAGGAGAT---GTTATATAACAAAGATTGTCCACCATACCTCATTT	760
Oy	298	GluAenleuYs-----LyleuGluValAenlysserAen	309
Db	761	GATGTAAAGAAAGATATATCCATGCGCAACCATACTGCGGTACTCGCTCATATTTCAAT	820
Oy	310	lysSerilleYsAsnpheGluAenArgPheleuPheArglyGlyGluValSerAsprrP	329
Db	821	CACCTCATCTCAAAATTT-----ATGAGAAAGAGATGCTCGAGACTGG	865
Oy	330	ValAenYrLeuSerProSergJinValGluArgleuSerAlaleuValAspAspYsleu	349
Db	866	AAGAACCCTTACTGCTGCGCTATGATGAGAACTTGTATGACATTATGAAAAAGAGATG	925
Oy	350	GlyGlySerGlyLeuThrPheArgleu	358
Db	926	GCAGGCTCCACACTGAACCTTGCCCTG	952

RESULT 6

US-10-199-334-3

Sequence 3, Application US/10199334

Patent No. 6905855

GENERAL INFORMATION:

APPLICANT: WOODAGE, Trevor et al.

TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING

TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN

TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF

FILE REFERENCE: CLO00669DIV-3

CURRENT APPLICATION NUMBER: US/10/199,334

CURRENT FILING DATE: 2002-07-22

PRIOR APPLICATION NUMBER: 09/609,816

PRIOR FILING DATE: 2001-03-27

PRIOR APPLICATION NUMBER: 60/192,408

PRIOR FILING DATE: 2000-03-27

PRIOR APPLICATION NUMBER: 60/212,725

PRIOR FILING DATE: 2000-06-20

PRIOR APPLICATION NUMBER: 09/609,816

PRIOR FILING DATE: 2000-07-03

NUMBER OF SEQ ID NOS: 23

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

LENGTH: 1004

TYPE: DNA

ORGANISM: Homo sapiens

US-10-199-334-3

Alignment Scores:

Pred. No.:	2,01e-26	Length:	1004
Score:	300.00	Matches:	95
Percent Similarity:	42.82%	Conservative:	63
Best Local Similarity:	25.75%	Mismatches:	121

Query Match:	15.81%	Indels:	90
DB:	3	Gaps:	14

  

US-10-019-931-3 (1-359) x US-10-199-334-3 (1-1004)	
QY	16 PheSerMetCysHisIleValLeuGluIleuLeuValGluGluIleValGlySerThrArgAspValProIys 35
Db	16 PheSerMetCysHisIleValLeuGluIleuLeuValGluGluIleValGlySerThrArgAspValProIys 35
QY	38 TTCGCCCTTAATGGCAGAAATTGAG-----AAAAACGCTCCACG 76
Db	38 TTCGCCCTTAATGGCAGAAATTGAG-----AAAAACGCTCCACG 76
QY	36 AlaGluGluAspGluGluIleuSerCysGluPheGluMet--LeuAspSerLeuPro 54
Db	36 AlaGluGluAspGluGluIleuSerCysGluPheGluMet--LeuAspSerLeuPro 54
QY	77 ATGGAAGAAAAAGCCGAACTG-----TTTAACTATCATGAAAGTAAATGAGATGCCCT 127
Db	77 ATGGAAGAAAAAGCCGAACTG-----TTTAACTATCATGAAAGTAAATGAGATGCCCT 127
QY	55 LysGluArgGlyTTPArgThrArgTyrIleuTyrIleuPheGluGlyPheThr----- 71
Db	55 LysGluArgGlyTTPArgThrArgTyrIleuTyrIleuPheGluGlyPheThr----- 71
QY	128 ACG-----TTGATATTATCAAAAGAAATGGGCGGAAAAAGTC 163
Db	128 ACG-----TTGATATTATCAAAAGAAATGGGCGGAAAAAGTC 163
QY	72 CysGlnAlaIalysGluIleGluAlaIleMetSerPheGluIleHisIlePheGlnSerLeuGlu 91
Db	72 CysGlnAlaIalysGluIleGluAlaIleMetSerPheGluIleHisIlePheGlnSerLeuGlu 91
QY	164 TGT-----AATTCCAGCCAAAGCCT 184
Db	164 TGT-----AATTCCAGCCAAAGCCT 184
QY	92 AsnAspValValLeuAlaIleThrIleProIysSerGlyThrThrTyrIleuValAlaLeuThr 111
Db	92 AsnAspValValLeuAlaIleThrIleProIysSerGlyThrThrTyrIleuValAlaLeuThr 111
QY	185 GATGATCTTATTTCTGGCAACTTACCAGAAAGTCAGGTAACAACGATGATGATAATTTTA 244
Db	185 GATGATCTTATTTCTGGCAACTTACCAGAAAGTCAGGTAACAACGATGATGATAATTTTA 244
QY	112 PheThrIleLeuAsn-----Arg 117
Db	112 PheThrIleLeuAsn-----Arg 117
QY	245 GACATGATTTCAAAATGATGTGATGTGAGAAATGCAAAAGAGCCGACCTAGATAGA 304
Db	245 GACATGATTTCAAAATGATGTGATGTGAGAAATGCAAAAGAGCCGACCTAGATAGA 304
QY	118 HisArgPheAspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHis 137
Db	118 HisArgPheAspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHis 137
QY	305 CACGGTTTC-----CTTGAAGCTAAATTTCCCAT 334
Db	305 CACGGTTTC-----CTTGAAGCTAAATTTCCCAT 334
QY	138 AspleuValProPhePheGluTyrIleValLeuTyrIalAsnGlyAspValProAspleuSer 157
Db	138 AspleuValProPhePheGluTyrIleValLeuTyrIalAsnGlyAspValProAspleuSer 157
QY	335 AAAGAAAAACGAGTTTGGAGTTGCTGCTT----- 364
Db	335 AAAGAAAAACGAGTTTGGAGTTGCTGCTT----- 364
QY	158 GlyLeuAlaSerProArgThrPheAlaThrHisIleuProPheGlySerIleuValLeuThr 177
Db	158 GlyLeuAlaSerProArgThrPheAlaThrHisIleuProPheGlySerIleuValLeuThr 177
QY	365 GAAATGTCCTCACCAACATGATAAAACACATCTCCCTTCACATCTGATTCACCCATCT 424
Db	365 GAAATGTCCTCACCAACATGATAAAACACATCTCCCTTCACATCTGATTCACCCATCT 424
QY	178 IleGluIysProGlyValIleValValIleValIleValIleValIleValIleValIleVal 197
Db	178 IleGluIysProGlyValIleValValIleValIleValIleValIleValIleValIleVal 197
QY	425 ATTCGAAAAAGAACTGCAGAGTTTGTCTATGTGGCCAGAAATCCAGAGATTGCCGTG 484
Db	425 ATTCGAAAAAGAACTGCAGAGTTTGTCTATGTGGCCAGAAATCCAGAGATTGCCGTG 484
QY	198 SerSerTrpHisTyrThrAsnAsnIleIysSerGluSerValSerProValIleuLeuAsp 217
Db	198 SerSerTrpHisTyrThrAsnAsnIleIysSerGluSerValSerProValIleuLeuAsp 217
QY	485 TTCCTACTACCACTTT--CACAGATGGCTCTTATGCCGATCTCTAGAACTTAGAG 541
Db	485 TTCCTACTACCACTTT--CACAGATGGCTCTTATGCCGATCTCTAGAACTTAGAG 541
QY	218 GlnAlaPheAspLeuTyrCysArgGlyValIleGlyPheGlyProPheTrpGluHisMet 237
Db	218 GlnAlaPheAspLeuTyrCysArgGlyValIleGlyPheGlyProPheTrpGluHisMet 237
QY	542 GAATTTATATGAGAAATTCATGTCGCGAAAGATTGTTGGCGGCTCCGATTGACCATGTG 601
Db	542 GAATTTATATGAGAAATTCATGTCGCGAAAGATTGTTGGCGGCTCCGATTGACCATGTG 601
QY	238 LeuGlyTyrTPArgGluSerIleuValArgProGluValIlePhePheLeuArgTyrGlu 257
Db	238 LeuGlyTyrTPArgGluSerIleuValArgProGluValIlePhePheLeuArgTyrGlu 257
QY	602 AAAGAGATGTGG-----GCTGCAGAAAGACACGACCGAGTCTCTACCTTCTACGAG 655
Db	602 AAAGAGATGTGG-----GCTGCAGAAAGACACGACCGAGTCTCTACCTTCTACGAG 655
QY	258 AspleuIysAspAspIleGluThrAsnIleuIysArgLeuAlaIleThrPheLeuGluIleuPro 277
Db	258 AspleuIysAspAspIleGluThrAsnIleuIysArgLeuAlaIleThrPheLeuGluIleuPro 277
QY	656 GATATTAAAAAAATCCAAAACATGAGATCCACAGAGGTGGAAATTTCTTGGAGAAAAC 715
Db	656 GATATTAAAAAAATCCAAAACATGAGATCCACAGAGGTGGAAATTTCTTGGAGAAAAC 715
QY	278 PheThrGluGluGluIleuArgIleGlyValValAlaIleAlaGluIleuCysSerPhe 297
Db	278 PheThrGluGluGluIleuArgIleGlyValValAlaIleAlaGluIleuCysSerPhe 297
QY	716 TTTCAGAGTAT-----GTTATTAACAAGATTGTCCACCATACCTCATTT 760
Db	716 TTTCAGAGTAT-----GTTATTAACAAGATTGTCCACCATACCTCATTT 760
QY	298 GluAsnLeuIys-----LysLeuGluValAlaAsnIysSerAsn 309
Db	298 GluAsnLeuIys-----LysLeuGluValAlaAsnIysSerAsn 309
QY	761 GATGATATGAAGAGTAATCCATGCGACCAACATACCTAGCGGTAACCTGCATCATATTCAAT 820
Db	761 GATGATATGAAGAGTAATCCATGCGACCAACATACCTAGCGGTAACCTGCATCATATTCAAT 820
QY	310 LysSerIleIysAsnPheGluAsnArgPheLeuPheArgIleGlyValIleValSerAspTrp 329
Db	310 LysSerIleIysAsnPheGluAsnArgPheLeuPheArgIleGlyValIleValSerAspTrp 329
QY	821 CACTTCATCTCAAAATTT-----ATGAGGAAAGGAGTGCCTGGAGACTGG 865
Db	821 CACTTCATCTCAAAATTT-----ATGAGGAAAGGAGTGCCTGGAGACTGG 865

Oy	330	ValentylrLeuSerPProSeInValGluAgluSeRAlaleuValAspharyleu	349
Db	866	AAGAACCCCTTACCTGGCGCTTGAAATGAAGAACTTTGATTAAAGCATTAATAAGAAAGATG	925
Oy	350	GlyGlySerGlyLeuthrPheargLeu	358
Db	926	GCAGGCTCACACTGAACCTTCGCTG	952
<b>RESULT 7</b>			
US-10-199-329-3			
; Sequence 3, Application US/10199329			
; Patent No. 6953681			
; GENERAL INFORMATION:			
; APPLICANT: WOODAGE, Trevor et al.			
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING			
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN			
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF			
; FILE REFERENCE: CLO00659DIV-2			
; CURRENT APPLICATION NUMBER: US/10/199,329			
; PRIOR FILING DATE: 2002-07-22			
; PRIOR APPLICATION NUMBER: 09/609, 816			
; PRIOR FILING DATE: 2001-03-27			
; PRIOR APPLICATION NUMBER: 60/192,408			
; PRIOR FILING DATE: 2000-03-27			
; PRIOR APPLICATION NUMBER: 60/212,725			
; PRIOR FILING DATE: 2000-06-20			
; PRIOR APPLICATION NUMBER: 09/609, 816			
; PRIOR FILING DATE: 2000-07-03			
; NUMBER OF SEQ ID NOS: 23			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 3			
; LENGTH: 1004			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-10-199-329-3			
<b>Alignment Scores:</b>			
Pred. No.:	2,01e-26	Length:	1004
Score:	300.00	Matches:	95
Percent Similarity:	42.82%	Conservative:	63
Best Local Similarity:	25.75%	Mismatches:	121
Query Match:	15.81%	Indels:	90
DB:	3	Gaps:	14
US-10-019-931-3 (1-359) x US-10-199-329-3 (1-1004)			
Oy	16	PhesermCyuhieyleuGluleuLeuLysglulYlthrArgrarValProlye	35
Db	38	TTCGCCCTTATGCGCAAGAATTGAG-----AAAAAGCCTCCACG	76
Oy	36	AlagIgluLaerGIglYleuSerCySGluPheGlnGluMet--LeuaPserLeuPro	54
Db	77	ATGCAAAAAGCCCAACTG-----TTTAACATCATGAAGTAGAGTAGTCCCT	127
Oy	55	LysgluArvgilyTrparGThrArGYleuTYleuPheGlnGluPheTr-----	71
Db	128	ACG-----TTGATATTATCAAAGAAATGTGGAAAAAGTC	163
Oy	72	CySGlnAlalyegIuleGlnAlalleMeSerPheGlnIynhiSPheGlnsertleuglu	91
Db	164	TGT-----AATTTCCAAAGCCAGCCT	184
Oy	92	AsnaerValaleuAlathrlileProLySerGlyThrtThrPleuLyAlaleuThr	111
Db	185	GATGATCTTATTTGGCAACTTACCAGAGTCAGGTACACAATGATGATGAATTTTA	244
Oy	112	PheThrIleLeuAn-----Arg	117
Db	245	GACATGATTTAAATGATGGTGAATGAGAAATGC AAAAGCCAGACTGTGATAGA	304
Oy	118	HieArGrPheArPrroValAlaserSetrThraShierPoleuPheThrSerAenProhis	133
Db	305	CACGCTTTC-----CTTGAACATAATTTCCCCT	334

```

QY 138 AspleuValProPhePheGluTyrLysLeuTyrAlaAsnGlyAspValProAspleuSer 157
Db 335 AAAGAAAAACCAAGTTTGAGTTCCTCTTT----- 364
QY 158 GlyLeuAlaSerProArgThrPheAlaThrHisLeuProPheGlySerLeuArgLutThr 177
Db 365 GAATGTCCTCCACCAACTGTATAAAAACATCTTCCTTCACATCTGATTCACACATCT 424
QY 178 IleGlyLysProGlyValIleValIleTyrLeuCyArgAsnProPheAspThrPheIle 197
Db 425 ATTCGAAAGAAAACTGCAGAGATTGTCTATGTGGCCAGAAATCCCAAGGATTGCCGTGG 484
QY 198 SerSerTrpHisTyrThrAsnAsnIleLysSerGluSerValSerProValLeuLeuAsp 217
Db 485 TCCTCACTACCACTTT--CACGAGATGGCTCTTTATGCCGTATCTCAGAACTTAGAG 541
QY 218 GluAlaPheAspleuTyrCyArgGlyValIleGlyPheGlyProPheTrpGluHisMet 237
Db 542 GAATTTATGGAATTCATGTCGCGAAAAAGTGTGGCGGCTCTGTTGACCATGG 601
QY 238 LeuGlyTyrTrpArgGluSerLeuLysArgProGluLysValPhePheLeuArgTyrGlu 257
Db 602 AAAGATGATGG-----GCTCAAAAGACACAGCACCGGATCCTCACTCTTCACGAG 655
QY 258 AspleuLysAspAspIleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuPro 277
Db 656 GATATTTAAAAAATCCAAACATGATGATCCCAAGGTTTGGAATTCCTTGGAGAAAACT 715
QY 278 PheThrGluGluGluLysArgLysGlyValIleValIleAlaGluLeuCySerPhe 297
Db 716 TTGTCAAGTGAT-----GTTATAAACAAGATTGTCCACCATACCTCATTT 760
QY 298 GluAsnLeuLys-----LysLeuGluValAsnLysSerAsn 309
Db 761 GATGTAATGAAGGATATATCCATGCGCAACCATACTCGGTACCTGGCTACATATTTCAAT 820
QY 310 LysSerIleLysAsnPheGluLeuAsnArgPheLeuPheArgLysGlyValIleSerAspTrp 329
Db 821 CACTCCATCTCAAAATTT-----ATGAGAAAGGAGTATGCTGAGACTGG 865
QY 330 ValAsnTyrLeuSerProSerGlnValGluArgLeuSerAlaLeuValAspAspLysLeu 349
Db 866 AAGAACCACCTTACTGTGGCTATGATGATGAACCTTGATTAAGCATATGAAAAAGAGATG 925
QY 350 GlyLysSerGlyLeuThrPheArgLeu 358
Db 926 GCAGGGTCACACCTGAACCTTGGCTGG 952

RESULT 8
US-09-609-816-1
/ Sequence 1, Application US/09609816
/ Patent No. 6436684
/ GENERAL INFORMATION:
/ APPLICANT: Woodage, Trevor
/ APPLICANT: Wei, Minh Hui
/ APPLICANT: Kodira, Chinnappa
/ APPLICANT: Beasley, Ellen
/ APPLICANT: DiFrancesco, Valentina
/ TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
/ TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
/ FILE REFERENCE: CL000669PCF
/ CURRENT APPLICATION NUMBER: US/09/609, 816
/ PRIOR FILING DATE: 2000-07-03
/ PRIOR APPLICATION NUMBER: 60/192, 408
/ PRIOR FILING DATE: 2000-03-27
/ PRIOR APPLICATION NUMBER: 60/212, 725
/ PRIOR FILING DATE: 2000-06-20
/ PRIOR APPLICATION NUMBER: 09/609, 816
/ PRIOR FILING DATE: 2000-07-03
/ NUMBER OF SEQ ID NOS: 23
/ SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 1
; LENGTH: 942
; TYPE: DNA
; ORGANISM: HUMAN
US-09-609-816-1

Alignment Scores:
Pred. No.: 4,83e-26 Length: 942
Score: 296.50 Matches: 81
Percent Similarity: 45.76% Conservative: 54
Best Local Similarity: 27.46% Mismatches: 105
Query Match: 15.62% Indels: 55
DB: 3 Gaps: 8

US-10-019-931-3 (1-359) x US-09-609-816-1 (1-942)

QY 86 HispGlnSerLeuGluAsnAspValValLeuAlaThrIleProlySerGlyThrThr 105
DB 128 AATTCCAAAGCCAGCGATGATCTTATTCGGCAACTTACCAAGTCAGGTACCA 187
QY 106 TrpLeuValAlaLeuThrPheThrIleLeuAsn----- 116
DB 188 TGGATGCATGAATTTTACACATGATCTTAATGATGGATGGAGAAATGCAAAAGA 247
QY 117 -----ArgHisArgPheAspProValAlaSerSerThrAsnHisProleu 131
DB 248 GCCCAGACTCTGATAGACACGCTTC-----CTT 277
QY 132 PheThrSerAsnProHisAspLeuValProPhePheGluTyrLeuLeuTyrAlaAsnGly 151
DB 278 GAATGAAATTTCCCATTAAGAAAACAGATTGGAGTTGCTT----- 325
QY 152 AspValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPhe 171
DB 326 -----GAAATGTCCTCACCAACCACTGATATAAAACACATCTCCCTTCA 367
QY 172 GlySerLeuValGluThrIleGluLysProGlyValValLeuValTyrLeuGlyAsn 191
DB 368 CATCTGATTTCCACCATCTTCTGGAAGAAAACCTGCMAATTTCTATGTGGCCGAAAT 427
QY 192 PropheAspThrPheIleSerSerTrpHisTyrThrAsnAsnIleLysSerGlySerVal 211
DB 428 CCCAAGATGTCCTGCTGCTCTACTACCACTTT---CACAGATGGCTTCTTATATGCT 484
QY 212 SerProValLeuLeuAspGlnAlaPheAspLeuTyrCysArgGlyValIleGlyPheGly 231
DB 485 GATCCTCAGAACTTGAAGAAATTTATGAAATTCATGTCGGAATACTTGTGGCGGG 544
QY 232 PropheTrpGluHisMetLeuGlyTyrTrpArgLysSerLeuLysArgProGluLysVal 251
DB 545 TCCTGCTTGAACCATGTGAAGAGATGGTGG-----GCTGCMAAAGACATGCACCGGATC 598
QY 252 PhePheLeuArgTyrGluAspLeuLysAspAspIleGluThrAsnLeuLysArgLeuAla 271
DB 599 CTCATCCTCTTCTACAGAGATATTAATAAAATCCAAACATGAGATCCACAGAGTGG 658
QY 272 ThrPheLeuGluLeuProPheThrGluGluGluValGlyValValIleGlyValIle 291
DB 659 GAATCTTGGGAAAACCTGTGTCAGGTGAT-----GTTATTAACAGAT 703
QY 292 AlaGluLeuCysSerPheGluAsnLeuLys-----LysLeu 303
DB 704 GTCCACCACTACTCTTATGATGTAATGAAGATTAATCCATGCGCAACCATCTGGGGTA 763
QY 304 GluValAsnLysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeuPheArgLys 323
DB 764 CCTGCTCACATATTCATCATCTCCATCTCAAAATTT-----ATGAGGAAA 808
QY 324 GlyGluValSerAspTrpValAsnTyrLeuSerProSerGlnValGluArgLeuSerAla 343
DB 809 GGGATGCCCTGAGATGGAAGAACCTTTACTGTGGCTTGAATGATGAACCTTTGATAG 868
QY 344 LeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArgLeu 358

```

```

DB 869 CATTAAGAAAGATGCGACGGTCCACACTGAACCTTGTGCTG 913
RESULT 9
US-10-199-334-1
; Sequence 1, Application US/10199334
; Patent No. 6905855
GENERAL INFORMATION:
APPLICANT: WOODAGE, Trevor et al.
TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
FILE REFERENCE: CLO00669DIV-3
CURRENT APPLICATION NUMBER: US/10/199,334
CURRENT FILING DATE: 2002-07-22
PRIOR APPLICATION NUMBER: 09/609,816
PRIOR FILING DATE: 2001-03-27
PRIOR APPLICATION NUMBER: 60/192,408
PRIOR FILING DATE: 2000-03-27
PRIOR APPLICATION NUMBER: 60/212,725
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/609,816
PRIOR FILING DATE: 2000-07-03
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 942
TYPE: DNA
ORGANISM: Homo sapiens
US-10-199-334-1

Alignment Scores:
Pred. No.: 4,83e-26 Length: 942
Score: 296.50 Matches: 81
Percent Similarity: 45.76% Conservative: 54
Best Local Similarity: 27.46% Mismatches: 105
Query Match: 15.62% Indels: 55
DB: 3 Gaps: 8

US-10-019-931-3 (1-359) x US-10-199-334-1 (1-942)

QY 86 HispGlnSerLeuGluAsnAspValValLeuAlaThrIleProlySerGlyThrThr 105
DB 128 AATTCCAAAGCCAGCGATGATCTTATTCGGCAACTTACCAAGTCAGGTACCA 187
QY 106 TrpLeuValAlaLeuThrPheThrIleLeuAsn----- 116
DB 188 TGGATGCATGAATTTTACACATGATCTTAATGATGGATGGAGAAATGCAAAAGA 247
QY 117 -----ArgHisArgPheAspProValAlaSerSerThrAsnHisProleu 131
DB 248 GCCCAGACTCTGATAGACACGCTTC-----CTT 277
QY 132 PheThrSerAsnProHisAspLeuValProPhePheGluTyrLeuLeuTyrAlaAsnGly 151
DB 278 GAATGAAATTTCCCATTAAGAAAACAGATTGGAGTTGCTT----- 325
QY 152 AspValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPhe 171
DB 326 -----GAAATGTCCTCACCAACCACTGATATAAAACACATCTCCCTTCA 367
QY 172 GlySerLeuValGluThrIleGluLysProGlyValValLeuValTyrLeuGlyAsn 191
DB 368 CATCTGATTTCCACCATCTTCTGGAAGAAAACCTGCMAATTTCTATGTGGCCGAAAT 427
QY 192 PropheAspThrPheIleSerSerTrpHisTyrThrAsnAsnIleLysSerGlySerVal 211
DB 428 CCCAAGATGTCCTGCTGCTCTACTACCACTTT---CACAGATGGCTTCTTATATGCT 484
QY 212 SerProValLeuLeuAspGlnAlaPheAspLeuTyrCysArgGlyValIleGlyPheGly 231
DB 485 GATCCTCAGAACTTGAAGAAATTTATGAAATTCATGTCGGAATACTTGTGGCGGG 544

```

```
QY 232 ProPheTrpGluH1m6etleuGlyTrTPArgGluSerleuLysArgProGluLysVal 251
DB 545 TCCTGGTTGACCATGTGAAGAGATGCG-----GTCGAAAGAACATGCACCGGATC 598
QY 252 PhePheLeuArgTyGluAspLeuLysAspAsp1leGluThrAsnleuLysArgLeuAla 271
DB 599 CTCCTACCTTTCTACAGAGATATATMAAAATCCAAAACATGAGATCCACAAAGGTGG 658
QY 272 ThrPheLeuGluLeuProPheThrGluGluGluArgLysGlyValValLysAlaIle 291
DB 659 GAATCTTGAGAAAACCTGGTCAGGTGAT-----GTTATMAACAAGATT 703
QY 292 AlaGluLeuCySeSerPheGluAsnleuLys-----LysLeu 303
DB 704 GTCCACCATACCTCATTTGATGTAATGAAGATATCCATGCCACCATCTACTGCGGTA 763
QY 304 GluValAsnLysSerAsnLysSer1leLysAsnPheGluAsnArgPheLeuPheArgLys 323
DB 764 CCGGCTCACATATTCAATCACTCATCTCAAAATTT-----ATGAGGAAA 808
QY 324 GlyGluValSerAspTrpValAsnTyTrLeuSerProSerGluValGluArgLeuSerAla 343
DB 809 GGGATGCTCGAGACTGGAAGAACCTTTACTGTGGCTTGAATGAGAACTTTGATTAAG 868
QY 344 LeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArgLeu 358
DB 869 CATTTAGAAAAGAAAGATGGCAGGGTCCACACTGAACCTTCGCTG 913

RESULT 10
US-10-199-329-1
; Sequence 1, Application US/10199329
; Patent No. 6953681
; GENERAL INFORMATION:
; APPLICANT: WOODAGE, Trevor et al.
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CLO00669DIV-2
; CURRENT APPLICATION NUMBER: US/10/199,329
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 942
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-199-329-1

Alignment Scores:
Pred. No.: 4,83e-26 Length: 942
Score: 296.50 Matches: 81
Percent Similarity: 45.76% Conservative: 54
Best Local Similarity: 27.46% Mismatches: 105
Query Match: 15.62% Indels: 55
DB: 3 Gaps: 8

US-10-019-931-3 (1-359) x US-10-199-329-1 (1-942)
QY 86 HisPheGlnSerLeuGluAsnAspValValLeuAlaThrIleProLysSerGlyThrThr 105
DB 128 AATTTCACACCAAGCTGATGATCTTATCTGCAACCTTAACCAAGATGAGTACACACA 167
QY 106 TrpLeuLysAlaLeuThrPheThrIleLeuAsn----- 116
DB 188 TGGATGATCAATAATTGTAGACATGATCTTAAGATGATGATGTGAGAAATGCAAAAGA 247
```

```
QY 117 -----ArgHisArgPheAspProValAlaSerSerThrAsnHisProLeu 131
DB 248 GCCCAGACTCTAGATAGACACCGCTTC-----CTT 277
QY 132 PheThrSerAsnProHisAspLeuValProPhePheGluTyTrLysLeuTyTrAlaAsnGly 151
DB 278 GAATGAAATTTCCCATATMAAGAAAACAGATTTGGAGTTGCTTCTT----- 325
QY 152 AspValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPhe 171
DB 326 -----GAAATGTCCTCACCAACATGATMAAAACATCTCCCTTCA 367
QY 172 GlySerLeuLysGluThrIleGluLysProGlyValLysValValTyTrLeuCyAsnAsn 191
DB 368 CATCTGATTCACCATCTATCTGGAAGAAAACTCGAAAGATGTCTATGTGGCCGAAAT 427
QY 192 ProPheAspThrPheIleSerSerTrpHisTyTrThrAsnAsn1leLysSerGluSerVal 211
DB 428 CCCAAGATTTGCTGGTGCTTACTACCACTTT---CACAGATGGCTTCTTTATGCGCT 484
QY 212 SerProValLeuLeuAspGlnAlaPheAspLeuTyTrCyAsnGlyValIleGlyPheGly 231
DB 485 GATCCTCAGAACTTGAAGAAATTTATGAAATTCATGTCGGAAGAACTGTGGCGGG 544
QY 232 ProPheTrpGluH1m6etleuGlyTrTPArgGluSerleuLysArgProGluLysVal 251
DB 545 TCCTGGTTGACCATGTGAAGAGATGCGT-----GTCGAAAGAACATGCACCGGATC 598
QY 252 PhePheLeuArgTyGluAspLeuLysAspAsp1leGluThrAsnleuLysArgLeuAla 271
DB 599 CTCCTACCTTTCTACAGAGATATATMAAAATCCAAAACATGAGATCCACAAAGGTGG 658
QY 272 ThrPheLeuGluLeuProPheThrGluGluGluArgLysGlyValValLysAlaIle 291
DB 659 GAATCTTGAGAAAACCTGGTCAGGTGAT-----GTTATMAACAAGATT 703
QY 292 AlaGluLeuCySeSerPheGluAsnleuLys-----LysLeu 303
DB 704 GTCCACCATACCTCATTTGATGTAATGAAGATATCCATGCCACCATCTACTGCGGTA 763
QY 304 GluValAsnLysSerAsnLysSer1leLysAsnPheGluAsnArgPheLeuPheArgLys 323
DB 764 CCGGCTCACATATTCAATCACTCATCTCAAAATTT-----ATGAGGAAA 808
QY 324 GlyGluValSerAspTrpValAsnTyTrLeuSerProSerGluValGluArgLeuSerAla 343
DB 809 GGGATGCTCGAGACTGGAAGAACCTTTACTGTGGCTTGAATGAGAACTTTGATTAAG 868
QY 344 LeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArgLeu 358
DB 869 CATTTAGAAAAGAAAGATGGCAGGGTCCACACTGAACCTTCGCTG 913

RESULT 11
US-09-328-174A-2
; Sequence 2, Application US/09328174A
; Patent No. 6448003
; GENERAL INFORMATION:
; APPLICANT: Guida, Marco
; APPLICANT: Kuitert, Janice
; TITLE OF INVENTION: Genotyping Human Phenol Sulfotransferase
; TITLE OF INVENTION: (STP2)
; FILE REFERENCE: 4389-6 (formerly SEQ-16P)
; CURRENT APPLICATION NUMBER: US/09/328,174A
; CURRENT FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: 09/328,174
; PRIOR FILING DATE: 1999-06-08
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1396
; TYPE: DNA
; ORGANISM: H. sapiens
```

```

; FEATRE:
; : NAME/KEY: CDS
; : LOCATION: (426) ... (1308)
US-09-328-174A-2

Alignment Scores:
Pred. No.:      1,81e-25      Length:      1396
Score:          294.00        Matches:       77
Percent Similarity: 47.48%    Conservative:  55
Best Local Similarity: 27.70% Mismatches:   150
Query Match:     15.49%      Indels:       26
DB:              3           Gaps:         8

US-10-019-931-3 (1-359) x US-09-328-174A-2 (1-1396)

QY PheGlnSerLeuGluAsnArgValValIleuAlaThrIleProLysSerGlyThrThrTrp 106
Db |||.....:|||||||.....:|||||||.....:|||||||.....:|||||||.....: 584
555 TTCACAGGGCCCGGCCTGATCACCCTGCTCATCAAGCACACTTACCCCAAGTCGGACACACTGG 584

QY LeuLysAlaLeuThrPheThrIleLeuAsnArgHisArgPheAspProValAlaSerSer 126
Db :::::.....:|||||||.....:|||||||.....:|||||||.....:|||||||.....: 635
585 GTGAGCCAGCATTTCTGGACATGATC-----TACCAAGGGGGGTGACCTGGAAAAAGTGT 635

QY ThrAsnHisProLeuPheThrSerAsnProHisAspLeuValProPhePheGlyTrlys 146
Db |||.....:|||||||.....:|||||||.....:|||||||.....:|||||||.....: 680
636 CACCGAGCTCCCATCTTCATGCGG-----GTGCCCTTCCTTGAGTTGAA 680

QY LeuTryAlaAsnGlyAspValProAsp-----LeuSerGlyLeuAlaSerPro 162
Db |||.....:|||||||.....:|||||||.....:|||||||.....:|||||||.....: 731
681 GTC-----CCAGGGATTCCCTCAGGAATGAGACTGTGAAAACAACACAGCCCCA 731

QY ArgThrPheAlaThrHisLeuProPheGlySerLeuLysGlyThrIleGluLysProGly 182
Db CGACCTCTGAAGAACAACACTGGCCCCCGGTCTGCTGCCACAGCTGTTGGATCAAGAG 791
722

QY ValLysValValIleuCyAspAsnProPheAspThrPheIleSerSerTriphistyr 202
Db |||.....:|||||||.....:|||||||.....:|||||||.....:|||||||.....: 851
792 GTCAAAGTGSTCTATGTTCCCGCAACGCAAGAGATGTGGCGTTCTTCACTACACTTC 851

QY ThrAsnAsnIleLysSerLysSerLysSerProValLeuLeuAspGlnIlaPheAspLeu 222
Db |||.....:|||||||.....:|||||||.....:|||||||.....:|||||||.....: 908
852 TACCAACATGCCCAAAATGTACCT---CACTCCGGAGCCTGGGAAAGCTTCTCGTAGAAG 908

QY TyrCyAspArgIValIleGlyPheGlyProPheTrpGlnHiseMetLeuGlyTyrrTParg 242
Db :::::|||||||.....:|||||||.....:|||||||.....:|||||||.....: 965
909 TTCATAGGCTGGAGAAGATGCTCTATGGTCTCTGTACACAGCACTGCAAGAAGTCTGG--- 965

QY GluSerLeuLysArgProGluLysValPhePheLeuArgTrgIuaAspLeuLysAspAsp 262
Db |||.....:|||||||.....:|||||||.....:|||||||.....:|||||||.....: 102
966 ---GAGCTAGGCCGCAACCACTGTTCTCTACTCTTCTATGAAACATGAAGAGAGAAC 102

QY IleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuProPheThrIleGluGlu 282
Db :::::|||||||.....:|||||||.....:|||||||.....:|||||||.....: 107
1023 CCCAAAAGGAGATTCMAAAGATCTCGAGTTGTGGGGCGCTCCTGCCAGAGGAG--- 107

QY GluArgLysGlyValValLysAlaIleAlaGluLeuCySerPheGluAsnLeuLysLys 302
Db |||.....:|||||||.....:|||||||.....:|||||||.....:|||||||.....: 112
1080 -----ACTGTGACCTCATAGTGTGGACACACGTCTTCAAGAGAGAAAGAG 112

QY LeuGluValAsnLysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeu----- 320
Db :::::|||||||.....:|||||||.....:|||||||.....:|||||||.....: 118
1128 AACCTTATGACCAACTACACACACCGTCCGCGGAGATTATGAGACCAAGACATCTCCCC 118

QY ---PheArgLysGlyGluValSerAspTrpValAsnTrylLeuSerProSerGlnValGlu 339
Db |||.....:|||||||.....:|||||||.....:|||||||.....:|||||||.....: 124
1188 TTCATAGAGAAAAGGCTGGCTGGGAGCTGGAGACCAACCTTCAACCTGGGCGAGATGAG 124

QY ArgLeuSerAlaLeuValAspAspLysLeuGlyGlySerGlyLeuThrPheArg 357
Db |||.....:|||||||.....:|||||||.....:|||||||.....:|||||||.....: 1301
1248 CGCTTCGATGCCGACTATGCGGAGAAAGATGGACGCTTCAGCTTCAGCTTCCC 1301

RESULT 12
US-09-949-016-4457

```

```

; Sequence 4457, Application US/09949016
; Patent No. 681239
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFECTION WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4457
; LENGTH: 1045
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4457

Alignment Scores:
Pred. No.:          9,35e-25          Length:          1045
Score:              266.50            Matches:          85
Percent Similarity: 47.33%            Conservative:     57
Best Local Similarity: 28.33%          Mismatches:      121
Query Match:        15.09%            Indels:          37
DB:                 3                  Gaps:            11

US-10-019-931-3 (1-359) x US-09-949-016-4457 (1-1045)

```

QY 75 LysGIuIIeGIaIIleMeSerPheGIn-----Lys 85  
 Db 140 GAAGAGGCGCATGGGATTTCTAATGATATAAGATTTTGTCAATATATTGGATTAATGTGGA 199  
 QY 86 HisPheGInSerLeuGIuIaSnAspValValLeuAlaThrILeProLysSerGIuYrThr 105  
 Db 200 GCGTTCCAGGCAAGACCAAGATATCTTGTCATGGCCACCTTACCCTTAATCTGGTCAACC 259  
 QY 106 TrPLeuYsAlaLeuThrPheThrILeUenAsnArgHisArgPheAspProValAlaSer 125  
 Db 260 TCGGTATTGAAGAAATGTGTATATGATC-----TATAAGAGGGGTATGTGGAAAG 310  
 QY 126 SerThrAsnHisProLeuPheThrSerAsnProHisAspLeuValProPhePheGIuYr 145  
 Db 311 TGCAAGAAGAGTATATTTTT-----AATCGAATCCCTTCCCGGAATGC 355  
 QY 146 LysLeu-----TyrAlaAsnGIuAspValProAspLeuSerGIuLeuAlaSerProArg 163  
 Db 356 AGAAAGAAGAAACCTCATGATGGA---GTAAACAATTGATGAGATGAATTTCTCCAGA 412  
 QY 164 ThrPheAlaThrHisLeuProPheGIuSerLeuYsGIuThrILeGIuLysProGIuVal 183  
 Db 413 ATGTGGAAGACATCTATTGGCACCTGAACCTTCTCCGCTCATTTTGGGAAAGAAATGT 472  
 QY 184 LysValValTyrLeuCysArgAsnProPheAspThrPheILeSerSerTrpHisYrThr 203  
 Db 473 AAGATATCTATCTTTGGCCGAATCAAGAGATGGGCTGTCTCTTATATATTTCTTT 532  
 QY 204 AsnAsnILeLysSerGIuSerValSerProValLeuLeuAspGlnAlaPheAspLeuYr 223  
 Db 533 CTAATGTGGCTGGTCATCCAA--AATCTGGATCTTTCCAGAGAGTTGTGGGAATTC 589  
 QY 224 CysArgGIuValILeGIuPheProPheTrpGlnHisMeLeuGIuYrTrpArgGIu 243  
 Db 590 ATGCACAGACAGAGTTCCTTATGTTCTCGGTATTAACAATGATAAATCTGGTGG---GAA 646  
 QY 244 SerLeuLysArgProGIuLysValPhePheLeuAspTyrGIuLysPheLysAspArgILe 263  
 Db 647 AAGGGAAGAAGTCCA---GCTGTATATTTCTTTCTACGAAGACCTGGAAGAGGATATC 703

QY 264 GIUThraenleuylargyleu1aThrPheleuGluleuProPheThrgluGluglu 283  
 Db 704 AGAAAAAGCATATAAATTGATACATTCTCGAAAAAGAACACAGAGAG----- 757  
 QY 284 ArglysglyValVallysa1a1lealagluLeuCySerPhegluaenleuylsleu 303  
 Db 758 -----CTGTGGACAGGATTATACATCACTTCCGTTCCAGAGATGAAGAAC-- 805  
 QY 304 GIUValaenlySerAsnlys-----SerIlelysaenPhegluaen 317  
 Db 806 -----AATCCATCCACAAATTACACACACCTGCCAGAGAAATTATGAACGAAATTGG 859  
 QY 318 ArgPheleuPheArglysglygluValSerAspTirpValaenlyrleuSerProsergin 337  
 Db 860 TCGCCCTTCATGAGAAAGGAATTACAGAGACTGGAAAAATCATCTTACAGTACGCCCTG 919  
 QY 338 ValGIUArgleuSerAla1eulVala1eapaby1sleuGlySerGlyleuThrPhearg 357  
 Db 920 AATGAAAAATTGATTAACATTATGACGACAAATGAAGAAATCTACACTGAAGTTTGA 979

## RESULT 13

US-08-325-562-1  
 / Sequence 1, Application US/08325562  
 / Patent No. 5714594  
 / GENERAL INFORMATION:  
 / APPLICANT: Aksoy, Ibrahim A.  
 / APPLICANT: Aksoy, Ibrahim A.  
 / APPLICANT: Wood, Thomas C.  
 / TITLE OF INVENTION: CDNA CLONING AND EXPRESSION OF HUMAN  
 / TITLE OF INVENTION: LIVER ESTROGEN SULFOTRANSFERASE  
 / NUMBER OF SEQUENCES: 17  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Schwesman, Lundberg & Woessner P.A.  
 / STREET: 3500 IDS Center  
 / CITY: Minneapolis  
 / STATE: MN  
 / COUNTRY: USA  
 / ZIP: 55402  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: Patentin Release #1.0, Version #1.25  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/325,562  
 / FILING DATE: 18-OCT-1994  
 / CLASSIFICATION: 435  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Muehling, Ann M.  
 / REGISTRATION NUMBER: 33,977  
 / REFERENCE/DOCKET NUMBER: 150.131US1  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: 612-339-0331  
 / TELEFAX: 612-339-3061  
 / INFORMATION FOR SEQ ID NO: 1:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 1063 base pairs  
 / TYPE: nucleic acid  
 / STRANDEDNESS: single  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: DNA  
 / FEATURE:  
 / NAME/KEY: CDS  
 / LOCATION: 107..989  
 / US-08-325-562-1

## Alignment Scores:

Pred. No.: 9,6e-25 Length: 1063  
 Score: 286.50 Matches: 85  
 Percent Similarity: 47.33% Conservative: 121  
 Best Local Similarity: 28.33% Mismatches: 37  
 Query Match: 15.09% Indels: 11  
 DB: 2 Gaps: 11

US-10-019-931-3 (1-359) x US-08-325-562-1 (1-1063)

QY 75 LysGIUllieGlnAla1leMetSerPhegin-----Lys 85  
 Db 140 GAAGAAAGTCCATGGGATCTTAATGTATTAAGATTTTGCAAATATTTGGATTAATGTGAA 199  
 QY 86 HisPheGlnSerLeuGluaenAspValValleu1aThr1leProlySerGlyThrThr 105  
 Db 200 GCGTTCCAGGAGAACGACATGATCTTGTCACTGCGACCTACCTTAATCTGGTCAACC 259  
 QY 106 TrpLeuVala1eulThrPheThr1leLeu5aenArgHisArgPheAspProValAlaSer 125  
 Db 260 TCGGTTAGTAAATTTGTATATATGATC-----TATAAAGAGCGTGATGTGAAAG 310  
 QY 126 SerThrAsnHisProLeuPheThSerAsnProHisAspLeuValProPhegluThr 145  
 Db 311 TGCAAAAGAAAGATGTAATTTT-----AATCGAATACCTTTCTCGAAATGC 355  
 QY 146 Lysleu-----TyrAlaasnGlyAspValProAspLeuSerGlyLeuAlaSerProArg 163  
 Db 356 AGAAAAAGAAACCTCATGATGGA---GTAAACAAATTAGATGATGATTAATCTCTGA 412  
 QY 164 ThrPheAlaThrHisLeuProPheGlySerLeuLysGIUThr1leGluLysProGlyVal 183  
 Db 413 ATTGGAAGACTCATTTGCGACCTGAACCTTCTCGCTCATTTTGGGAAAGATTGT 472  
 QY 164 LysValVal1TyrluCyArgAsnProPheAspThrPheIleSerSerTriPhe1aThr 203  
 Db 473 AAGATAACTCATCTTTCGCGAATGCAAAGATGCGCTGTTCTTTTATATTTCTTT 532  
 QY 204 Asnaen1lelySerGluSerValSerProValleuLeuAspGlnAlaPheAspLeuThr 223  
 Db 533 CTAAATGGTGGCTGCGATCCA---AATCTGGATCTTTCAGAGTTGTGAGAAATTC 589  
 QY 224 CyArgGlyVal1leGlyPheGlyProPheTrpGluHisMetLeuGlyTyrTrpArgGlu 243  
 Db 590 ATGCAGAGACAGGTTCTTATGTGTTCTCGATTAACATGTAAATCTTGGTG--GAA 646  
 QY 244 SerLeuLysArgProGluLysValaPhePheLeuArgTyrGluAspLeuLysAsp1le 263  
 Db 647 AAGGAAAGAGTCC---CGTGTACTATTTCTTTCTCGAAGACCTGAAGACAGATATC 703  
 QY 264 GIUThraenleuylargyleu1aThrPheleuGluleuProPheThrgluGluglu 283  
 Db 704 AGAAAAAGCATATAAATTGATACATTCTCGAAAAAGAACACAGAGAG----- 757  
 QY 284 ArglysglyValVallysa1a1lealagluLeuCySerPhegluaenleuylsleu 303  
 Db 758 -----CTGTGGACAGGATTATACATCACTTCCGTTCCAGAGATGAAGAAC-- 805  
 QY 304 GIUValaenlySerAsnlys-----SerIlelysaenPhegluaen 317  
 Db 806 -----AATCCATCCACAAATTACACACACCTGCCAGAGAAATTATGAACGAAATTGG 859  
 QY 318 ArgPheleuPheArglysglygluValSerAspTirpValaenlyrleuSerProsergin 337  
 Db 860 TCGCCCTTCATGAGAAAGGAATTACAGAGACTGGAAAAATCATCTTACAGTACGCCCTG 919  
 QY 338 ValGIUArgleuSerAla1eulVala1eapaby1sleuGlySerGlyleuThrPhearg 357  
 Db 920 AATGAAAAATTGATTAACATTATGACGACAAATGAAGAAATCTACACTGAAGTTTGA 979

## RESULT 14

US-08-437-795-1  
 / Sequence 1, Application US/08437795  
 / Patent No. 574335  
 / GENERAL INFORMATION:  
 / APPLICANT: Aksoy, Ibrahim A.  
 / APPLICANT: Aksoy, Ibrahim A.  
 / APPLICANT: Wood, Thomas C.  
 / TITLE OF INVENTION: CDNA CLONING AND EXPRESSION OF HUMAN  
 / TITLE OF INVENTION: LIVER ESTROGEN SULFOTRANSFERASE



NUMBER OF SEQUENCES: 19  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Schwegman, Lundberg & Woessner P.A.  
 STREET: 3500 IDS Center  
 CITY: Minneapolis  
 STATE: MN  
 COUNTRY: USA  
 ZIP: 55402

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/437,795  
 FILING DATE: 09-MAY-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mueeling, Ann M.  
 REGISTRATION NUMBER: 33,977  
 REFERENCE/DOCKET NUMBER: 150.137US1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 612-339-0331  
 TELEFAX: 612-339-3061  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1063 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 107..989  
 US-08-437-795-1

Alignment Scores:  
 Pred. No.: 9.6e-25 Length: 1063  
 Score: 286.50 Matches: 85  
 Percent Similarity: 47.33% Conservative: 57  
 Best Local Similarity: 28.33% Mismatches: 121  
 Query Match: 15.09% Indels: 37  
 Gaps: 11

US-10-019-931-3 (1-359) x US-08-437-795-1 (1-1063)

75 Lysglulleglnalailemetserpneqln-----lys 85  
 140 GAAGAAGTCATGGGATTTCTAATGATAAAGTTTTCACAAATATTGGGATTAATGTGAA 199  
 86 Hispneqlnsrleuqluasnspvalyleuathrlleprolyserglythrthr 105  
 200 GCGTCCGAGGACAGACCATGATCTTGTCAATCCCTCAATTCGTGACCAACC 259  
 106 Trpleuysalaleuthrphethrilleuasnarghisargpneaspvalalaser 125  
 260 TGGGTACTGAATTTGTATATGATC-----TATAAAGGGGTGATGAGAAAG 310  
 126 Serthraenhisproleuuphetrserasnprohisaspleuvalprophspheglutyr 145  
 311 TGCAGAAAGATGTAATTTT-----AATCGAATACCTTTCTCTGGAATGC 355  
 146 Lysleu-----tyralasnglyaspvalproaspleuyserglyleualaserproarg 163  
 356 AGAAAAAGAAAACCATGATATGGA---GTAAAAACAATTAGATGATGATATCTCTGAGA 412  
 164 Thrthraenhisleuysleuuphetheglyserleuysglutthrlleuglyserproglyval 183  
 413 ATTGTGAAGACTCATTTTCCACCTGAACCTTCTCTGCTCAATTTTGGGAAAAGGATGCT 472  
 184 LysvalvaltyrleucyargaspneaspheapthrpheilesersertrpHisertythr 203  
 473 AAGATTAATCTATCTTTTGGCGGAATGCAAAAGATGTGGCTTTCTTTATATATTTCTTT 532

204 Asnanellelysserqluservalserprovalleuasnarglnalaphaspleuetyr 223  
 533 CTAATGGTGGCTGTCAATCCA--AAATCTGATCTTTCCAGAGTTTGGAGAAATTC 589  
 224 Cysargglyvalilleglypneqlprophetrpqlunhismetleuglytyrtrparglu 243  
 590 ATGCAAGACACAGTCTCTTATGTTCTCTGTAATAACATGTAATCTGTGG---GAA 646  
 244 Serleuysargprogluysvalpnepheleuargtyrgluaspleuysaspaspille 263  
 647 AAGGAAAAGAGTCCA---CGGTACTATTTCTTTCTACGAAGACCTGAAAGAGATATC 703  
 264 Glutthraenleuysargleuvalaleuthrpheluileuuprophethnugluuglu 283  
 704 AGAAAAGAGTGAATTAATGATCATTTCTCTGAAAGAAACCATTCAGAGAG--- 757  
 284 Arglysglyvalvalalealagluuencysserpneqluasnleuylsleu 303  
 758 -----CTGTGACACAGATTATACATCATCTTCTTCCAAGAGATGAAGAAC--- 805  
 304 Gluvalenlysserasnlys-----Serllelyasnpheglun 317  
 806 -----AATCATTCACAAATTAACAACACTGCCAGACGAAATTAATGAAACGAAATTG 859  
 318 ArgpheleuphearglysglygluvalseraspthrpvalaenlyrleuSerproserqln 337  
 860 TCGCCCTTCATGAAGAGGAAATTACAGAGACTGAGAAATCACTTTACAGTACGCCCTG 919  
 338 Valgluargleuseralaleuvalaspsaplyseuqlglyserglyleuthrphearg 357  
 920 AATGAAAAATTTGATTAACATTAATGACGACGAATGAAGATCTACACTGAAGTTTGA 979

RESULT 15  
 US-09-717-321A-45  
 ; Sequence 45, Application US/09717321A  
 ; Patent No. 6852845  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gould-Rothberg, Bonnie E.  
 ; APPLICANT: DiPippo, Vincent A.  
 ; APPLICANT: Ramsen, Tenore M.  
 ; APPLICANT: Gerwein, Robert W.  
 ; TITLE OF INVENTION: METHOD OF IDENTIFYING TOXIC AGENTS USING NSAID-INDUCED  
 ; FILE REFERENCE: 15966-601 Utility  
 ; CURRENT FILING DATE: 2000-11-20  
 ; PRIOR APPLICATION NUMBER: 60/166,923  
 ; PRIOR FILING DATE: 1999-11-22  
 ; NUMBER OF SEQ ID NOS: 50  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 45  
 ; LENGTH: 993  
 ; TYPE: DNA  
 ; ORGANISM: Rattus norvegicus  
 US-09-717-321A-45

Alignment Scores:  
 Pred. No.: 8.09e-24 Length: 993  
 Score: 278.50 Matches: 86  
 Percent Similarity: 47.12% Conservative: 61  
 Best Local Similarity: 27.56% Mismatches: 121  
 Query Match: 14.67% Indels: 45  
 Gaps: 14

US-10-019-931-3 (1-359) x US-09-717-321A-45 (1-993)

57 Argglytyrptargthraargtyrleu----- 66  
 4 AAGGCTGGAATACTAAAG-TTATCATGATGTCAAGCTATACTTGTGTTGAAGAAATA 62  
 67 ---PheglnglypnetpCygnalalyseglulleglnalilemetserpneqlnys 85

```
Db      63 COTTTTCGCGCTTTGGTT---TCAAGAAATT-----CTGAAATAATAGTTGAAG 113
Qy      86 H1pRheGlnSerLeuGluAsnAspValValLeuAlaThrIleProIysSerGlyThrThr 105
Db      114 AAGTTTGTGTAAGAAAGACGATTCATGATTTGACTTACCCCAAGTCAGAAACGAAAC 173
Qy      106 TrpLeuLysAlaLeuThrPheThrIleLeuAsnArgHisArgPheAspPro-----Val 123
Db      174 TGGCTG-----ATCGAATTGTCTGCTTGATTCAGACCAAGGAGATCCCAAGTGATC 227
Qy      124 AlaSerSerThrAsnHisProLeuPheThrSerAsnProHisAspLeuValProPhePhe 143
Db      228 CAATCTATG-----CCCATCTG-----GATCGCTCACCCCTGATA 263
Qy      144 GluTyrLysLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAlaSerProArg 163
Db      264 GAG-----ACTGTTCAAGATATGATTAATTAAACAAATGGAAGACACGAA 311
Qy      164 ThrPheAlaThrHisLeuProPheGlySerLeuLysGluThrIleGluLysProGlyVal 183
Db      312 CTGATGACCTCCCATCTTCCATGCATCTTTCTCCAAAGTCTCTTCAGTTCCAAGGCC 371
Qy      184 LysValValTyrLeuCysArgAsnProPheAspThrPheIleSerSerTrpHisTyrThr 203
Db      372 AAGGTGATATATCTCATCGAAATCCAGAGATGTTCTGTTCCTGCTTAATTTTCTCG 431
Qy      204 AsnAsnIleLysSerGluSerValSerProValLeuAspGlnAlaPheAspLeuTyr 223
Db      432 AGTAAGATCGCCCTGAGAGAG--AAACCACTCGCTGGAACTTACGTTAAATGGTTC 488
Qy      224 CysArgGlyValIleGlyPheGlyProPheTrpGluHisMetLeuGlyTyrTrpArgGlu 243
Db      489 CTCAAAGGAAATGTTGCATATGATCATGCTTGAGCATCATCGTGGCTGGCTG----- 542
Qy      244 SerLeuLysArgProGluLysValPhePheLeuArgTyrGluAspLeuLysAspAspIle 263
Db      543 TCTATGAGAGAAATGGACAACTTCTTGTACTGTACTATGAAGACATGAAAAAGCATACA 602
Qy      264 GluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuProPheThrGluGluGlu 283
Db      603 ATGGATCCATTAAGAAAGATATGTGACTTCTGGGAAAAAAATTAGAGCCAGATGAG-- 659
Qy      284 ArgLysGlyValValIleLysAlaIleAlaGluLeuCysSerPheGluAsnLeuLysLysLeu 303
Db      660 -----CTGAATTGCTCCTCAAGTATAGTTCCTTCCAAAGTCGTGAAA----- 701
Qy      304 GluValAsnLysSerAsnLysSerIleLysAsnPheGlu-----AsnArgPheLeu 320
Db      702 GAAAAACAACATGCTCAATATAGCCCTCATGAGAGAAAGAACTGATTCCTAGGTTTACT 761
Qy      321 Phe---ArgLysGlyGluValSerAspTrpValAsnTyrLeuSerProSerGlnValGlu 339
Db      762 TTCATGAGAAAAAGGCAACTAATGCTGAAAGAAATCACTTCACAGTACCCCAAGCTGAA 821
Qy      340 ArgLeuSerAlaLeuValAspAspLysLeuGlyGly 351
Db      822 GCCTTGATTAAGTGTTCACAGAGAAATGGCCGGT 857
```

Search completed: December 23, 2005, 21:46:15  
Job time : 226 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame\_p2n model

Run on: December 23, 2005, 19:08:24 ; Search time 834 Seconds  
(without alignments)  
3559.600 Million cell updates/sec

Title: US-10-019-931-3

Perfect score: 1898  
Sequence: 1 MATSSMKSIMAIPFSFMCH.....RLSALVDDKLGSGSLFTFRLS 359

Scoring table:  
BIOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 9793542 segs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=framed\_p2n.model -DEV=x1p  
-Q=/cgn2\_1/USPRO.epool.p/US10019931/runat.16122005.165220.18500/app\_query.fasta\_1.519  
-DB=Published Applications NA\_Main -QFMT=fastcap -SUFFIX=p2n.rnpbm  
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1  
-MATRIX=blonum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext  
-HEAPSIZE=500 -MILEN=0 -MAXLEN=2000000000  
-USBR=US10019931.OCN 1.1549 @runat.16122005.165220.18500 -ICPU=6 -ICPU=3  
-NO\_MMAP -LARGESOURCY -NEG\_SCORES=0 -WAIT -DSPBLCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications NA\_Main:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1058.5	55.8	1311	7	US-10-424-599-58871
2	770.5	40.6	996	3	US-09-938-842A-1027
3	770.5	40.6	996	3	US-09-938-842A-1027
4	763.5	39.2	1069	8	US-10-767-795-558
5	757	39.9	981	3	US-09-938-842A-2049
6	757	39.9	981	3	US-09-938-842A-2049
7	725.5	38.2	787	7	US-10-424-599-82242

8	674	35.5	1041	7	US-10-437-963-55857	Sequence 55857, A
9	672	35.4	1440	8	US-10-767-795-1749	Sequence 1749, Ap
10	671.5	35.4	271990	6	US-10-195-144-87	Sequence 87, Appl
11	671.5	35.4	271990	6	US-10-345-072-87	Sequence 87, Appl
12	666.5	35.1	1251	8	US-10-767-795-1750	Sequence 1750, Ap
13	666	35.1	1358	8	US-10-739-930-1663	Sequence 1663, Ap
14	659.5	34.7	1077	8	US-10-437-963-55524	Sequence 55524, A
15	648.5	34.2	1579	8	US-10-425-115-167844	Sequence 167844, A
16	648.5	34.2	1579	8	US-10-437-963-31855	Sequence 31855, A
17	635	33.5	1021	7	US-10-437-963-94656	Sequence 94656, A
18	633	33.4	1068	7	US-10-437-963-94656	Sequence 94656, A
19	614	32.3	1050	6	US-10-259-194A-15	Sequence 15, Appl
20	614	32.3	1050	6	US-10-437-963-54999	Sequence 54999, A
21	613.5	32.3	2067	6	US-10-195-144-80	Sequence 80, Appl
22	613.5	32.3	2067	6	US-10-345-072-80	Sequence 80, Appl
23	610	32.1	1410	8	US-10-425-115-101459	Sequence 101459, A
24	609	32.1	1366	7	US-10-767-701-10838	Sequence 10838, A
25	601.5	31.7	1011	7	US-10-437-963-96500	Sequence 96500, A
26	598	31.5	1017	6	US-10-259-194A-3	Sequence 3, Appl1
27	598	31.4	1017	7	US-10-437-963-57879	Sequence 27879, A
28	595.5	31.4	1056	7	US-10-437-963-51349	Sequence 51349, A
29	592	31.2	1403	7	US-10-425-114-21478	Sequence 21478, A
30	591.5	31.2	1526	7	US-10-425-114-27795	Sequence 27795, A
31	591.5	31.2	1461	6	US-10-259-194A-131	Sequence 131, Appl
32	589.5	31.1	1482	7	US-10-260-238-1003	Sequence 1003, Ap
33	589.5	31.1	1482	7	US-10-437-963-75177	Sequence 75177, A
34	589	31.0	984	6	US-10-259-165-443	Sequence 443, Appl
35	588.5	31.0	987	6	US-10-259-165-443	Sequence 113, Appl
36	588.5	31.0	1032	6	US-10-259-194A-329	Sequence 329, Appl
37	586.5	30.9	1041	6	US-10-437-963-56891	Sequence 56891, A
38	586.5	30.9	1041	7	US-10-259-165-531	Sequence 331, Appl
39	583	30.7	1828	8	US-10-437-963-68828	Sequence 68828, A
40	550.5	29.0	1192	3	US-10-425-115-168303	Sequence 168303, A
41	546	28.8	1033	7	US-10-437-963-72606	Sequence 72606, A
42	545.5	28.7	1435	7	US-10-437-963-101470	Sequence 101470, A
43	529.5	27.9	1044	7	US-10-437-963-99158	Sequence 99158, A
44	521.5	27.5	1878	7	US-10-437-963-72617	Sequence 72617, A
45	520.5	27.4	1019	7	US-10-767-701-15199	Sequence 15199, A

#### ALIGNMENTS

RESULT 1  
US-10-424-599-58871  
Sequence 58871, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(513223)B  
CURRENT APPLICATION NUMBER: US/10/424, 599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 58871  
LENGTH: 1311  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(1311)  
OTHER INFORMATION: unsure at all n locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_24172C.1  
US-10-424-599-58871  
Alignment Scores:  
Pred. No.: 5,22e-119 Length: 1311  
Score: 1058.50 Matches: 191  
Percent Similarity: 74.71% Conservative: 66

Best Local Similarity:	55.53%	Mismatches:	83
Query Match:	55.77%	Indels:	4
DB:	7	Gaps:	2

QY	19	CyHhsluylsleuGlulLeuLeuLys - GluGluYlyThrTrgAspValProLysValaGluGlu	38
DB	53	TGTCATCATGCTTCCAGAAAGAAATGAATCCGAGAAAGGGGAGGAAATATACATATGAAGA	112
QY	38	uAspGluGlyLeuSerCyGluPheGluGluMetLeuAspSerLeuProLysGluArgGlu	58
DB	113	AGACAG- - - CTAACTCAAGAATGTAAGAGAGTGTATCTCTCTCTCTAGGAGAGAGG	169
QY	58	YTPArgThrArgIlyLeuYIlyLeuPheGluGlyPheTrpCyGlnAlaLysGluIleGlu	78
DB	170	TTGGAGAACACCTTATATATCTATCTTTTCAAGAAATTTGGTGCCAGCCATTGGAAATCCA	229
QY	78	nalalIleMetSerPheGluLysPheHisPheGluSerLeuGluAsnAspValValIleuAlaTh	98
DB	230	AGCAATTAATCATTTTTCAGAGACATTTCCAGCTCAAGACAGTGAATGTTATTTGCGCAC	289
QY	98	rIleProLysSerGlyThrThrTrpLeuLysAlaLeuThrPheThrIleLeuAsnArgHi	118
DB	290	AATTCCAAAATCAGGTATCACCTTGCTGAAGCTCTCACCTTTGGCCATTGTCCATCGCCA	349
QY	118	AsrPheAspProValAlaSerSerThrAsnHisProLeuPheTrpSerAsnProHisAs	138
DB	350	TACTCATAGTATCATATCATCATCATCATCATCTTTGGCTTACTTCTTAATCCCAAGA	409
QY	138	pLeuValProPhePheGluYIlyLeuLysIleValAsn- - - - - GluAspValProAsp	156
DB	410	ACTGTGCTCTTTCACATAGAAATACACCGTTTATGCTAAATGCCCTTACCATGTTCCAAACT	469
QY	156	uSerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPheGlySerLeuLysGlu	176
DB	470	ATCCAAATCATGACTGAGCCAAAGCTTTTGTGTACATATTCATTCATTCATGATGGCCAA	529
QY	176	uThrIleGluLysProGlyValIlyLeuValIlyLeuLysArgAsnAspProPheAspThrPh	196
DB	530	GTCCATCAAGAGATCCAAATAGTAAGATATTTATATATATGAGAAACCACTTGACACTTT	589
QY	196	eIleSerSerTrpHisIlyTrpAsnAsnIleLysSerGluSerValSerProValIleuLe	216
DB	590	TGTGCTCATGCTGATTTTCTCCACAAATTAAGCCAGACACTTACCGTAATTTGAAGCT	649
QY	216	uAspGlnAlaPheAspLeuYIlyCyAsrGlyValIleGlyPheGlyProPheTrpGluHi	236
DB	650	AGGGCAACCTTTTGAAAGATGTGCAAGAGATTAATAGGCTTTGTGTCCAACTGGAGCA	709
QY	236	aMetLeuGlyYIlyTrpArgIlySerLeuLysArgProGluLysValPhePheLeuArgTy	256
DB	710	AATGTTGGATTTATGGAAGAGATATACCTTAGCGCTAGTAAAGGTTTGTGTTCTTGAAGTA	769
QY	256	rGluAspLeuLysAspAspIleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLe	276
DB	770	CGAGGATCTTAAATAAAGATGTCATATTTCAATGTGAAGAAATAGCGAGTCTTAGATG	829
QY	276	uProPheThrGluGluGluGluLysArgLysGlyValValIlyLysAlaIleAlaGluLeuCySe	296
DB	830	TCTCTTTCACCTTCGGAGGAAGAGGTGATGGGACTTATAGAGCAATATCAAGCTATGGAG	889
QY	296	rPheGluAsnLeuLysLysLeuGluValAsnLysSerAsnLysSerIleLysAsnPheGlu	316
DB	890	CTTCGAGAAGATGAAGAAATTTGGAAGCCAAATTAATCTGGAACATTTGCTAGAACCTTTGA	949
QY	316	uAsnArgPheLeuPheArgLysGlyGluValSerAspTrpValAsnTyIlyLeuSerProSe	336
DB	950	GAGAAAGTACTGTTCCGAAAGGCTGAATGGAGATTTGGGAGCACTACCTTCCCTGTA	1009
QY	336	rGluValGluArgLeuSerAlaLeuValAspAspLysLeuGluLysSerGlyLeuThrPh	356
DB	1010	AATGGGTAAAGATTATCGCAATATTATGGAAGAAAGTTAATGGTGTAGAGCTGTCTATT	1069

```
QY      356 eArgLeuSer 359
          |      ::|||
Db      1070 TTAAGTGTCT 1079
```

```

      2      US-09-938-842A-1027
      3      / Sequence 1027, Application US/093938842A
      4      / Patent No. US20020160378A1
      5      / GENERAL INFORMATION:
      6      / APPLICANT: Harper, Jeff
      7      / APPLICANT: Kieps, Joel
      8      / APPLICANT: Wang, Xun
      9      / APPLICANT: Zhu, Tong
     10      / TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
     11      / TITLE OF INVENTION: SAME, AND METHODS OF USE
     12      / FILE REFERENCE: SCRIPI300-3
     13      / CURRENT APPLICATION NUMBER: US/09/938,842A
     14      / CURRENT FILING DATE: 2001-08-24
     15      / PRIOR APPLICATION NUMBER: US 60/227,866
     16      / PRIOR FILING DATE: 2000-08-24
     17      / PRIOR APPLICATION NUMBER: US 60/264,647
     18      / PRIOR FILING DATE: 2001-01-16
     19      / PRIOR APPLICATION NUMBER: US 60/300,111
     20      / PRIOR FILING DATE: 2001-06-22
     21      / NUMBER OF SEQ ID NOS: 5379
     22      / SEQ ID NO 1027
     23      / LENGTH: 996
     24      / TYPE: DNA
     25      / ORGANISM: Arabidopsis thaliana
     26      / US-09-938-842A-1027

```

Pred. No.:	8,94e-84	Length:	996
Score:	770.50	Matches:	140
Percent Similarity:	62.50%	Conservative:	70
Best Local Similarity:	41.67%	Mismatches:	121
Query Match:	40.60%	Indels:	5
DB:	3	Gaps:	3

  

US-10-019-931-3 (1-359) x US-09-938-842A-1027 (1-996)	
Oy	22 LeuGluLeuLeuLeuGluGluGlyLeuThrArgArgValProIyValaGluGluArgGluGly 41
Db	1 ATGAAGGCTTCATGAAGA-----GCTATCACTCTCCAAATCATATGAAGAAGCAGAAC 54
Oy	42 IeuSerCysGluPheGluGluIleuLeuArgSerIeuProIyGluArgGlyTTPArgThr 61
Db	55 GTTAGTCAGAAACCAAGAACTTGATCACTCTTCACTTCAGACAAAGATTTCATGGGT 114
Oy	62 ArgTyrIleuTyrIleuPheGluGluIleuPheTyrCysGluAlaIyGluIleGluAlaIleMet 81
Db	115 TATGCTCTCTACAACTCAAAAGGTGGTGATCATATCCAAACACACTCCCAAGCGCTTCT 174
Oy	82 SerPheGluIyHisPheGluSerIeuGluIyGluIyValaIleuAlaThrIleProIyLe 100
Oy	175 GACGTCCAAAAACACTCAAGCAGAAATCTGTATATATCCCTGCTCTTGGCCAAA 234
Oy	102 SerGlyThrThrTrioIyLeuIyAlaIeuThrPheThrIleIleuAenArgHisArgPheAsp 121
Db	235 GGTGGAAACCACTTGCGTCAATCCCTAATTTTCGTGTGTGACATAGAAAAGTACCGC 294
Oy	122 ProValAlaSerSerThrAsnHisPheProIeuPheThrSerAsnProHisAspIleuValPro 141
Db	295 -----GGAACCCCTCAACACATCTTGTGCTTACAAAACCTCATGACCTTGCCCA 348
Oy	142 PhePheGluTyrIyIySerIeuTyrAlaIeuGluIyAspValProAspIleuSerGlyIleuAlaSer 161
Db	349 TTTCCTGAGAGTTGAGATTATACCTATATGCAAAATTCGCGATTCGCAAAAGTATTTCTTC 408
Oy	162 ProArgThrPheAlaThrHisIeuProIheGlySerIeuIyGluThrIleGluIyPro 181
Db	409 CCTATGATCTTTCTTATACACATGCACTTACCAAGATTTGCGTGAAGCCACACAAA--- 465

Qy	182	GlyValLysValValTyrIleuCyAhrGanPProPheAapTrPheIleSeSerTrpHis	201
Db	466	GCTTGCAAAACCGTATATGTGTGTAGAGGTATCAAAAGATACGTTGTCTCCGGCTGGCAT	525
Qy	202	TyrThrAsnAsnIleLysSerGluSerValSerProValLeuLeuAapGlnAlaPheAsp	221
Db	526	TATAGAAACATGTTGCATTCGCACCAAGATGATCAAGCACTTTGAGCTCATGTTTGT	585
Qy	222	LeuTyrCyAhrGlyValIleGlyPheGlyProPheTrpGlnHisMetLeuGlyTyrTrp	241
Db	586	GCATTATGTAGAGAACTTCTTTATATGCACTTATGGAAACATGATTTAGCTATTTGG	645
Qy	242	ArgGluSerIleuLysArgProGluLysValPhePheLeuArgTyrGlyAapLeuLysAap	261
Db	646	AAAGCGAGCTTGGAACCAAGAGAAATGTTCTTTTCATGAAGTACGAAGAAATATTAGC	705
Qy	262	AspIleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuProPheThrGluGlu	281
Db	706	GACCTCGTGTTCACACTCAAGAGACTCCGCCAGTCTTGGAAATGCTCATTCACCAAGAA	765
Qy	282	GluGluLysArgLysGlyValValLysAlaIleAlaGluLeuCySerPheGluAsnLeuLys	301
Db	766	GAAAGAAAGAAAGTGAATCCGTTGGAGAGATCTTGAACCTGTGTATTTACAAATTTAAGC	825
Qy	302	LysLeuGluValAsnLysSerAsnLysSerIleLysAsnPheGluAsnArgPheLeuPhe	321
Db	826	AATTGGAGGTTATATAGAAATGGACAACGAAATTTGTTGTAATTTCCACAGTCTTCTTT	885
Qy	322	ArgLysGlyGluValSerAspTrpValAsnTyrIleuSerProSerGlnValGluArgLeu	341
Db	886	AGGAAAGGTGAAGTTGGTATTTGAAAGATCATCTTACGCCACAATAGCGAAAACTTT	945
Qy	342	SerAlaLeuValAsnAspLysLeuGlyLysSerGlyLeuThrPheArg	357
Db	946	GATGACATTAATGACTATAGACTAGAGACTCCGGTTTGTATTTTCA	993

```
/ GENERAL INFORMATION:
/ APPLICANT: Kovacic, David K.
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Zhou, Yina
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
/ FILE REFERENCE: 38-21(53534)B
/ CURRENT APPLICATION NUMBER: US/10/767,795
/ CURRENT FILING DATE: 2004-01-30
/ NUMBER OF SEQ ID NOS: 117596
/ SEQ ID NO 558
/ LENGTH: 1069
/ TYPE: DNA
/ ORGANISM: Gossypium hirsutum
/ FEATURE:
/ OTHER INFORMATION: Clone ID: GOSHI-09MAY01-CL121_1
/ US-10-767-795-558

Alignment Scores:
Pred. No.: 7,28e-83 Length: 1069
Score: 763.50 Matches: 139
Percent Similarity: 64.17% Conservative: 67
Best Local Similarity: 43.30% Mismatches: 110
Query Match: 40.23% Indels: 5
DB: Gaps: 2

US-10-019-931-3 (1-359) x US-10-767-795-558 (1-1069)

QY 12 AAlaIleProSerPheSerMetCyshIleYbLeuGluLeuLeuGluGlyIsthrArg 31
DB 119 TCATTACCGAGCTGAAATGATTGGCCAT-----ACTCAAGCTGAAATTCAGTTTCT 169
QY 32 AspValProValAlaGluGluAspGluLeuSerCygluPheGlnGluMetLeuAsp 51
DB 170 GATGGTCCCAAAATACCTGCTAGATTGGGCTCACCACCCCAAGATGCAAGAGCTTATATCC 229
QY 52 SerLeuProLyseGluArgGlyTyrPargThrArgTyrLeuTyrLeuPheGlnGlyPheTyr 71
DB 230 TCACGTCCAGCGGAGAAAGGTTGGGTGGCCATCTCTCCATCAATATCAGGGGTTTGG 289
QY 72 CyshAlaLyseGluIleGlnAlaIleMetSerPheGlnLyshIshPheGlnSerLeuGlu 91
DB 290 CACACCACCTCCGAGTTTACACAGAGCTTACTTGGCAAAACACATTCACAGCTCAGAG 349
QY 92 AaaAspValValLeuAlaThrIlePheProLyseGlyThrThrTyrLeuValAlaLeuThr 111
DB 350 ACAAGATCTCTCTGTTACCACTCCAAATACGAGAAACATGGTTAAAGGCTATGTT 409
QY 112 PheThrIleLeuAsnArgHisArgPheAspProValAlaSerSerThrAsnIshProLeu 131
DB 410 TTGGCTTGATGAACGAGTCAAGTATCCC-----AACACGTATACATCACCCCTTGG 463
QY 132 PheThrSerAsnProHisAspLeuValProPhePheGluTyrLeuLeuTyrAlaAsnGly 151
DB 464 CTCTCAGAGAACCCCTCACATCTTGTTCATTTTGGAGTATGGGCTTACATGATAGT 523
QY 152 AspValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrIshIshProPhe 171
DB 524 CAGGTTCTCACTTCAACCACTTTTACATCTCCAGGCTTATGGAGATCATTTACCCCTT 583
QY 172 GlySerLeuLyseGluThrIleGluYbProGlyValIlyValIyTyrLeuCyArgAsn 191
DB 584 GTTTCATTCGCCAAGATCGGCAAAAGATCATCTTGCAAGCTTGTATTATATGAGAAC 643
QY 132 PropheAspThrPheIleSerSerTyrPheTyrAsnAlaIleLyseSerGluSerVal 211
DB 644 CCAAAAGATACCTTCTATCGCTTGGCACTTCCACCAACAGCTGAGAACCAAGATTAG 703
QY 212 SerProValLeuLeuAspGlnAlaPheAspLeuTyrCyArgGlyValIleGlyPheGly 231
DB 704 GGAAGCACTCTCTTGAGAGACCTTCGATTAAGTATACCAAGAGCGTGGATTGATGGA 763
QY 232 PropheTyrGluIshMetLeuGlyTyrTyrPargGluSerLeuLyshArgProGluYbVal 251
```

```
DB 764 CCGTTTGGACACATGTTTATGATGGAACAAACCTTGAAACCCGTGAAGAGCTC 823
QY 252 PhePheLeuArgTyrGluAspLeuLyshAspIleGluThrAsnLeuLyshArgLeuAla 271
DB 824 TTGTTTGAATATGAGAAATGAAAGAGCCACCACTTCACTAGCTATGAAAGTACT 883
QY 272 ThrPheLeuGluLeuProPheThrGluGluGluArgLybGlyValIyValIyAlaIle 291
DB 884 CAGTTCTCTCGATGCGCATTTTCCATGAGAGAGACTAGATGCGGTGATGATTA 943
QY 292 AlaGluLeuCySerPheGluAsnLeuLyshIshPheGlnGlyValIshIshSerAsnLyse 311
DB 944 CAAAACATGACAGCTTGAAGATTTGACCAACTGACGTTTAAACAATACCCGAAAGTTG 1003
QY 312 IleLyshAsnPheGluAsnArgPheLeuPheArgLybGlyGluValSerAspTyrValsh 331
DB 1004 GCATCGGCTGAGATTAACAGCATTTTTCACACACGTAATTTGAGATCTTAACAT 1063
QY 332 Tyr 332
DB 1064 CAT 1066

RESULT 5
US-09-938-842A-2049
/ Sequence 2049, Application US/09938842A
/ Patent No. US20020160378A1
/ GENERAL INFORMATION:
/ APPLICANT: Harper, Jeff
/ APPLICANT: Kreps, Joel
/ APPLICANT: Wang, Xun
/ APPLICANT: Zhu, Tong
/ TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
/ TITLE OF INVENTION: SAME, AND METHODS OF USE
/ FILE REFERENCE: S01P300-3
/ CURRENT APPLICATION NUMBER: US/09/938, 842A
/ PRIOR FILING DATE: 2001-08-24
/ PRIOR APPLICATION NUMBER: US 60/227,866
/ PRIOR FILING DATE: 2000-08-24
/ PRIOR APPLICATION NUMBER: US 60/264,647
/ PRIOR FILING DATE: 2001-01-16
/ PRIOR APPLICATION NUMBER: US 60/300,111
/ PRIOR FILING DATE: 2001-06-22
/ NUMBER OF SEQ ID NOS: 5379
/ SEQ ID NO 2049
/ LENGTH: 981
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
/ US-09-938-842A-2049

Alignment Scores:
Pred. No.: 3,99e-82 Length: 981
Score: 757.00 Matches: 148
Percent Similarity: 61.11% Conservative: 50
Best Local Similarity: 45.68% Mismatches: 120
Query Match: 39.88% Indels: 6
DB: Gaps: 3

US-10-019-931-3 (1-359) x US-09-938-842A-2049 (1-981)

QY 33 ValProLybAlaGluGluAspGluGlyLeuSerCygluPheGlnGluMetLeuAspSer 52
DB 19 GTTCTGCTTACTTGGAGATGAGATCTGCACAAAGAAACAAGAGCTGATCTTCTCT 78
QY 53 LeuProLybGluArgGlyTyrPargThrArgTyrLeuTyrLeuPheGlnGlyPheTyrCy 72
DB 79 CTTCCTAAAGAAAGGTTGTTAGTGAAGTAAATATATGATTCACAGACTTGGCAGC 138
QY 73 GlnAlaLyseGluIleGlnAlaIleMetSerPheGlnLyshIshPheGlnSerLeuGlu 92
DB 139 ACAACAGCTATTTTCAAGAAATCTTGTCTGCCAAABACCTTTGAAGCTTAAGATTC 198
QY 93 AspValValLeuAlaThrIlePheProLyseGlyThrThrTyrLeuValAlaLeuThrPhe 112
```

```

Db 199 GACATTATCCCGCTCACTAATCCTAATCAAGTACCACTTGTTAAAGCTTGCTTT 258
Qy 113 Thr1LeuAnaArgHisArgPheAspProValAlaSerSerThrAsnHisProLeuPhe 132
Db 259 GCTCTCCCTTAACCAACAAGATT---CCAGTTTCTTCTTGTAACCAATCCCTCTCTG 315
Qy 133 ThSerAnProHisAspLeuValProPhePheGluTyrIleuValAlaAsnGlyAsp 152
Db 316 GTACCAATCAACACCTTCTTGCTCCCTCTTGGAAGAGTTACTACGAGTCCCAAGAT 375
Qy 153 ValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPheGly 172
Db 376 TTC---GATTTCCTCAGTTTCCCTCTTCCAAAGACGTAGAACACGACATATCCACTT 432
Qy 173 SerLeuArgGluThrIleGluLysProGlyValIleLysValValIleLysCysArgAsnPro 192
Db 433 TCGCTCCCGAGTGTGTAAGAGCTCGTCTTGTAAGATTGTGTAATGTTGTAAGAACCT 492
Qy 193 PheAspThrPheIleSerSerThrHisTyrThrAsnAsnIleLysSerGluSerValSer 212
Db 493 AAGGACATGTTGTGCTTAATGGCATTTGGGAAAAAGCTAGCTCCTGAGAAAAACCGCG 552
Qy 213 ProValLeuAnaArgGlnAlaPheAspLeuTyrCysArgGlyValIleGlyPheGlyPro 232
Db 553 GATTATCTATCGAAAAAGCGGTGAAGCCTTTGTGAGGAAAGTTATATAGTGCACCC 612
Qy 233 PheTrpGluHisMetLeuGlyTyrTrpArgGluSerLeuLysArgProGluLysValPhe 252
Db 613 TTTTGGATCATATATATGAGACTAGTATCAAGCCGCAAGATCCGAAACAGATCTTGG 672
Qy 253 PheLeuArgTyrGluAspLeuLysAspAspIleGluThrAsnLeuLysArgLeuAlaThr 272
Db 673 TTTGTACTTACGAGAGCTTAAAGACAGACCAAGTTGAGATGAAGCGGATGCGGAG 732
Qy 273 PheLeuGluLeuProPheThrGluGluGluLysValIleValAlaIleAla 292
Db 733 TTTCTGGAATGGCTTATTTGAAGAAAGAA-----GTGAGAGATTTGTG 780
Qy 293 GluLeuCysSerPheGluAsnLeuLysLeuGluValAlaAsnLysSerAsnLysSerIle 312
Db 781 AAGTTGTAGCTTTGAGAGTTTAAGTAATTTGAGAGTTTAAAGAGGAAATTTGCCA 840
Qy 313 LysAsnHegLysAnaArgPheLeuPheArgLysGlyValAlaSerAspTrpValAsnTyr 332
Db 841 AATGAAATAGAGACTAAATCTTCTTGAAGAAAGAGATGAGAGATGAGATGACT 900
Qy 333 LeuSerProSerGlnValAlaArgLeuSerAlaLeuValAspAspLysLeuGlySer 352
Db 901 TTGAGTGAATGATTCGACAGAGAAATTCATGAACCATTTGAAGAAAGCTTTAAAGTTCT 960
Qy 353 GlyLeuThrPhe 356
Db 961 GGTCTTAATTT 972

```

## RESULT 6

```

US-09-938-842A-2049
; Sequence 2049, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16

```

```

; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2049
; LENGTH: 981
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2049

Alignment Scores:
Pred. No.: 3,99e-82 Length: 981
Score: 757.00 Matches: 148
Best Similarity: 61.11% Conservative: 50
Best Local Similarity: 45.68% Mismatches: 120
Query Match: 39.88% Indels: 6
Gaps: 3

US-10-019-931-3 (1-359) x US-09-938-842A-2049 (1-981)
Qy 33 ValProLysAlaGluGluAspGluGlyLeuSerCysGluPheGluMetLeuAspSer 52
Db 19 GTTCCTGCTTACTTGGAGATGAAAGATCTGACACAGAAACAAAGCTTGATCTTCT 78
Qy 53 LeuProLysGluArgGlyTyrTrpArgTyrLeuTyrLeuPheGlnGlyPheTrpCys 72
Db 79 CTTCCTAAAGAGAAAGCTTGTTAGTACGAAATATATGATTTCCAAAGACTTGGCAC 136
Qy 73 GlnAlaLysGluIleGlnAlaIleMetSerPheGluLysHisPheGlnSerLeuGluAsn 92
Db 139 ACACAGACTTTTATACAGAGAACTTGATCTGCCAAAAAGCTTGAAGCTTAAGATTCC 198
Qy 93 AspValValLeuAlaThrIleProLysSerGlyThrThrProLeuLysValAlaLeuThrPhe 112
Db 199 GACATTATCCCTGCTCACTAATCCTAATCAAGTACCACTTGTTAAAGCTTTGTCTTT 258
Qy 113 Thr1LeuAnaArgHisArgPheAspProValAlaSerSerThrAsnHisProLeuPhe 132
Db 259 GCTCTCCCTTAACCAACAAGATT---CCAGTTTCTTCTTGTAACCAATCCCTCTCTG 315
Qy 133 ThSerAnProHisAspLeuValProPhePheGluTyrIleuValAlaAsnGlyAsp 152
Db 316 GTACCAATCAACACCTTCTTGCTCCCTCTTGGAAGAGTTACTACGAGTCCCAAGAT 375
Qy 153 ValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPheGly 172
Db 376 TTC---GATTTCCTCAGTTTCCCTCTTCCAAAGCTAGTAACACGACATATCCACTT 432
Qy 173 SerLeuArgGluThrIleGluLysProGlyValIleLysValValIleLysCysArgAsnPro 192
Db 433 TCGCTCCCGAGTGTGTAAGAGCTCGTCTTGTAAGATTGTGTAATGTTGTAAGAACCT 492
Qy 193 PheAspThrPheIleSerSerThrHisTyrThrAsnAsnIleLysSerGluSerValSer 212
Db 493 AAGGACATGTTGTGCTTAATGGCATTTTGGGAAAAAGCTAGCTCCTGAGAAAAACCGCG 552
Qy 213 ProValLeuAnaArgGlnAlaPheAspLeuTyrCysArgGlyValIleGlyPheGlyPro 232
Db 553 GATTATCTATCGAAAAAGCGGTGAAGCCTTTGTGAAGGAAAGTTATATAGTGCACCC 612
Qy 233 PheTrpGluHisMetLeuGlyTyrTrpArgGluSerLeuLysArgProGluLysValPhe 252
Db 613 TTTTGGATCATATATATGAGACTAGTATCAAGCCGCAAGATCCGAAACAGATCTTGG 672
Qy 253 PheLeuArgTyrGluAspLeuLysAspAspIleGluThrAsnLeuLysArgLeuAlaThr 272
Db 673 TTTGTACTTACGAGAGCTTAAAGACAGACCAAGTTGAGATGAAGCGGATGCGGAG 732
Qy 273 PheLeuGluLeuProPheThrGluGluGluLysValIleValAlaIleAla 292
Db 733 TTTCTGGAATGGCTTATTTGAAGAAAGAA-----GTGAGAGATTTGTG 780
Qy 293 GluLeuCysSerPheGluAsnLeuLysLeuGluValAlaAsnLysSerAsnLysSerIle 312

```

Db	781	AAAGTTGCTGACCTTTGAGACGTTTAAAGTAATTTGGAAGCTTAACAAAGAAAGGAAATTTGCCA	840
QY	313	LYSAsnPhcgluAenArgrPheLeupPheArgrIysglYgluValSerAepTTPValAsnTyr	332
Db	841	AATGGAATAGAGACTAAACATTTCTTTTGAAAAGAGAGATTGGAGGATGGAGATACT	900
QY	333	LeuSerProSerGrInuAlGluArgrLeuSerAlaLeuValAspApyrLeuGlyIysSer	352
Db	901	TTGAGTGAGTCAATTTGGCAGAGAAATTGATAGAACATTGAAGAAAGTTTAAAGGTTCT	960
QY	353	GlyLeuThrPhe 356	
Db	961	GGTCTTAATTT 972	
RESULT 7			
US-10-424-599-82242			
; Sequence 82242, Application US/10424599			
; Publication No. US20040031072A1			
GENERAL INFORMATION:			
APPLICANT: La Roca Thomas J			
APPLICANT: Kovacic David K			
APPLICANT: Zhou Yihua			
APPLICANT: Cao Yongwei			
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With			
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement			
FILE REFERENCE: 38-21(53223)B			
CURRENT APPLICATION NUMBER: US/10/424, 599			
NUMBER FILING DATE: 2003-04-28			
NUMBER OF SEQ ID NOS: 285684			
SEQ ID NO 82242			
LENGTH: 787			
TYPE: DNA			
ORGANISM: Glycine max			
FEATURE:			
OTHER INFORMATION: Clone ID: PAT_MKT3847_45281C.1			
US-10-424-599-82242			
Alignment Scores:			
Pred. No.: 2,1e-78 Length: 787			
Score: 725.50 Matches: 131			
Percent Similarity: 72.98% Conservative: 50			
Best Local Similarity: 52.82% Mismatches: 62			
Query Match: 38.22% Indels: 5			
Gaps: 2			
US-10-019-931-3 (1-359) x US-10-424-599-82242 (1-787)			
QY	35	LYSAlaGluGluAepRGLuGlyLeuSerCysGluPheGIngluMetLeuAapSerLeuPro	54
Db	49	GAAAGCACTGAAAGAAACAAATTTAAAGCCAAATTTGTAAGAGATTATTCATCTTCTCT	108
QY	55	LYSGluArgrGlyTTPArgrThrArgrTyrLeuTyrLeuPheGInglYpHeTTPCysGlnAla	74
Db	109	AGAGAGAGGGTTGGATGAACACCGTATCTATTATTCACAAAGCTTTTGGTACTCATCA	168
QY	75	LYSGluIleGlnAlaIleMetSerPheGInglYshISpHeGInserLeuGluAAspVal	94
Db	169	ACTGAAATCCAAAGCCATTAACACTTTTCAAAACCAATTCCAAGCTTAAGGACAAATGATTT	228
QY	95	ValLeuAlaThrIleProLYsSerGlyThrThrTTPLeuLYsAlaLeuThrPheThrIle	114
Db	229	GTTATTTAGTTCAGTTCCCAAAATTCGGGCACCACTTGTTGAAAGCCCTTACTTTGGCATTT	288
QY	115	LeuAenArgrHISArgrPheAapProValAlaSerSerThrAenHISProLeuPheThrSer	134
Db	289	CTCCATCCGCCAGTATTTT-----CCTTCCTTAGAGAACCAACCCATTACTTATTTTTC	339
QY	135	AsnProHISaapLeuValProPhePheGluTyrLYsLeuTyr-----AlaAsnGlyAsp	152
Db	340	AATCCCACTGAACCTTGCTCCATCTTAATTAATTTGATTAAGTAATTAATTAATGGCCAA	399
QY	153	ValProAapRLeuSerGlyLeuAlaSerProArgrThrPheAlaThrHISLeuProPheGly	172

Db 400 AACTATGACCTATCCAAATATGCCAGCCAAAGATTTTGGGACTGATGTTCACTTCACT 459

Qy 173 SerLeuLySGluThrIleGluLySProGlyValLySValIyLeuCyAArgAenPro 192

Db 460 TCATTGGGCAATCATATTAAAGAGACTGACTGTAAGATCATTTATTTGTGTAGAAATCTA 519

Qy 193 PheAPThrPheIleSerSerTrpHisTyThrAsnAsnIleLySergIuSerValSer 212

Db 520 TTGACACCTTGTGTTCTTCACTTGGGGCTTTTGTGCACAAATTTTGCCAAAGTTTGCCCT 579

Qy 213 ProValLeuLeuAspGlnAlaPheAspLeuTyTCyAArgGlyValIleGlyPheGlyPro 232

Db 560 ACATTACCTCTTAGAGAGAACTTTTGAAAGATACGTGGAAGGGATTAATTTGGCTTGGTCCA 639

Qy 233 PheTrpGluHisMetLeuGlyTyTrpArgGluSerLeuLyAArgProGluLySValPhe 252

Db 640 TCTTGGAACCATATCTGATATTATGGAAAGAGACATTCCTAGGCCGAAGAAAGTTTGG 699

Qy 253 PheLeuArgTyArgGluAspLeuLySAspAlIleGluThrAsnLeuLySArgLeuAlaThr 272

Db 700 TTCTTGAAGATAGAGACCTTTAAAGAAATGTTCAATTATATGTGAAAGAAATAGCTGAG 759

Qy 273 PheLeuGluLeuProPheThrGlu 280

Db 760 TTCTTGGGTGTGCTTTCACATAA 783

RESULT 8

US-10-437-963-55857

/ Sequence 55857, Application US/10437963

/ Publication No. US20040123343A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

APPLICANT: Wu, Wei

APPLICANT: Boukharov, Andrey A.

APPLICANT: Boukharov, Brad

APPLICANT: Li, Ping

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53221)B

CURRENT APPLICATION NUMBER: US/10/437,963

CURRENT FILING DATE: 2003-05-14

NUMBER OF SEQ ID NOS: 204966

SEQ ID NO 55857

LENGTH: 1041

TYPE: DNA

ORGANISM: *Oryza sativa*

FEATURE:

OTHER INFORMATION: Clone ID: PAT\_MRT4530\_57824C.1

US-10-437-963-55857

Alignment Scores:

Pred. No.: 7,12e-72 Length: 1041

Score: 674.00 Matches: 139

Percent Similarity: 57.10% Conservative: 66

Best Local Similarity: 38.72% Mismatches: 130

Query Match: 35.51% Indels: 24

DB: 7 Gaps: 6

US-10-019-931-3 (1-359) x US-10-437-963-55857 (1-1041)

Qy 1 MetalathrSerMetLySerIleProMetAlaIleProSerPheSerMetCyAHis 20

Db 1 ATGCTTCTCTCTCCGTGCMAACCTCTCC-----TCGCCCTGCAT 42

Qy 21 LysLeuGluLeuLeuLySgluGlySThrArgAspValProLySAlaGluAspGlu 40

Db 43 GAAGTCGAC-----GCCGAAACTGACGAG 66

Qy 41 GlyLeuSerCyGluPheGluMetLeuAspSerLeuProLySgluArgGlyTrpArg 60



```

Db      67 GAACCTACAAGACGATTCACCGAGTCCTGTCGCGGAGCTCAGAACCATGCGCC 126
Qy      61 ThrArgTyrLeuTyrLeuPhe---GlnGlyPheTyrCysGlnAlaIleGluAla 79
Db      127 TTCCTCCCGCTCTACCGCCAGCAGCAAGGCTGTACAGTACCTCATGCTATGTCGT 186
Qy      80 IleMetSerPheGlnIlePheGlnSerLeuGluAlaMetPheValIleLeuAlaThrIle 99
Db      187 GCGATGTCGCCGACGCGCGGTTGCGCGCGCCCTCCGACATCATGTGTCGCACTTGG 246
Qy      100 ProLysSerGlyTyrThrTyrPheLeuValAlaLeuThrPheThrIleLeuAlaArg 119
Db      247 CCCAAGTCCCGCAGCAGCGTATCAAGCGCTCTCTACGCCAGCGATGACACCGG---AGG 303
Qy      120 PheAspProValAlaSerSerThrAsnHisProLeuPheThrSerAsnProHisAspLeu 139
Db      304 GAGCACCCTCCGCGACGCGCGCGCCGACCACTTTCATCTCCGCGCCCGCAGAGTGC 363
Qy      140 ValProPhePheGluTyrTyrLeuLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeu 159
Db      364 GTCAACTTCTCTCGAGTACAGCTCTACCAACCAACAGAGTCCGAGCTTGGCAGGCTTG 423
Qy      160 AlaSerProArgThrPheAlaThrHisLeuProPheGlySerLeuGlyLeuThrIleGlu 179
Db      424 CCGGACCCGAGGCTGTCGCGACGAGTCCGTTCACTGCTGCGGACGCGCGCGCG 483
Qy      180 LysProGlyValIleValValIleTyrLeuGlyAsnAspProPheAspThrPheIleSerSer 199
Db      484 GCGTCGGGCTGCAAGTGTGTACGTGTGCGCGACCCCAAGCAACCTGATCTCGATG 543
Qy      200 TrpHisTyrThrAsnAlaIleLysSer---GluSerValSerProValLeuLeuAspGln 218
Db      544 TGGACCTTCGCCCAACATTCAAGGCGCGGTAGAGGCGCAGAGCCCATATCCCGGAGGCC 603
Qy      219 AlaPheAspLeuTyrCysArgGlyValIleGlyPheGlyProPheTyrGluHisMetLeu 238
Db      604 ATGCCCGAGCTTCTGCTGCTGCGGCTGTGCGCGTCCGCGCGTACTGGACCAAGCTCTC 663
Qy      239 GlyTyrTyrPheArgLysSerLeuLysArgProGluLysValPhePheLeuArgTyrGluAsp 258
Db      664 GGCCTACGCGGCGCGCGCAGTGTGCGCGCCCGCAGAGTCTCTTCTTCAAGGTACAGAG 723
Qy      259 LeuLysAspAspIleGluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeuProPhe 278
Db      724 ATGAAAGCTCGAGCGCGCGCGCGCGCGCTCGAGAGCTGTGCGGCTGCGCTTCC 783
Qy      279 ThrGluGluGluGluArgLysGlyValIleValIleAlaIleAlaGluLeuCysSerPheGlu 298
Db      784 AGCGCGGAGGAGGAGGAAGGTGGCGGTGTGAGACGCGCATGTCAAGCTGTGCTCGTTGAT 843
Qy      299 AsnLeuLysLysLeuGluValAlaAsnLysSerLysSer-----IleLysAsnPhe 315
Db      844 CACATGATCGCGCTGTGAGCGCACCAAGAGCGCAGAGCTGTGTGTCGCGCAGCGCG 903
Qy      316 GluAsnArgPheLeuPheArgLysGlyValIleAspTyrPheValAsnTyrLeuSerPro 335
Db      904 GCGAATGACTCATCTTCTGCGCGCGGAGGTGGCGGCTGTGCGCAACATCTTTCGCGG 963
Qy      336 SerGluValGluArgLeuSerAlaLeuValAspAspLysLeuGlyLysSerGlyLeu 354
Db      964 GAGATTGACAGCGGATCGACGCATACCGAGCGCAGGTTCAATGTTCCGGTTC 1020

```

RESULT 9  
US-10-767-795-1749

; Sequence 1749, Application US/10767795  
; Publication No. US20040181830A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Zhou, Yihua  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement  
; FILE REFERENCE: 38-21(53534)B

```

; CURRENT APPLICATION NUMBER: US/10767,795
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO: 1749
; LENGTH: 1440
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C40_3
US-10-767-795-1749

Alignment Scores:
Pred. No.: 2,12e-71 Length: 1440
Score: 672.00 Matches: 136
Percent Similarity: 58.31% Conservative: 71
Best Local Similarity: 38.31% Mismatches: 122
Query Match: 35.41% Indels: 26
DB: Gaps: 8

US-10-019-931-3 (1-359) x US-10-767-795-1749 (1-1440)
Qy      8 SerIleProMetAlaIleProSerPheSerMetCysHisLysLeuGluLeuLeuLysGlu 27
Db      155 AATATGATTTTGCATCTATGCTGCTGCTC-----ATGGAATCCCATCTTGAG 205
Qy      28 GlySerThrArgAspValProLysAlaGluGluAspGluGlyLeuSerCysGluPheGln 47
Db      206 AAACAATAATGAGATGTG-----CTTCAAAATCTTTCAAA 241
Qy      48 GluMetLeuAspSerLeuProLysGluArgGlyTyrArg---ThrArgTyrLeuTyrLeu 66
Db      242 GAGATGATTTCTACTTCCCTTAAGAGGTTGTTGGGTTTCCGAAGATCAATATCA 301
Qy      67 PheGlnGlyPheTyrCysGlnAlaLysGluIleGlnAlaIleMetSerPheGlnLysHis 86
Db      302 TATCAAAATTTTGTGTTTCAGCCCTTTTCTACAGAGGATGTGCTGCTCAACAA 361
Qy      87 PheGlnSerLeuGluAsnAspValIleLeuAlaThrIleProLysSerGlyThrThrTyr 106
Db      362 TTCAGAGCTCAGCCCACTGATATATCTTGTGACTCTTCCAAAGACGACAGCCCTGG 421
Qy      107 LeuLysAlaLeuThrPheThrIleLeuAsnArgHisArgPheAspProValAlaSerSer 126
Db      422 TTTAAATTCCTCATTTTCCACATTTACAGAACTTCATAC-----AATGAT 469
Qy      127 ThrAsnHisProLeuPheThrSerAsnProHisAspLeuValProPhePheGluTyrLys 146
Db      470 TCCACCAACCCCTTACTTCTTCCAGATGCTCATGATGTGTGCTTACATGAGATTGAT 529
Qy      147 LeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAlaSerProArgThrPheAla 166
Db      530 -----CATGCCAGATTTCATATGACATCTTGAATTCCT---CTTTTACGC 577
Qy      167 ThrHisLeuProPheGlySerLeuLysGluThrIleGluLysProGlyValIleValIle 186
Db      578 ACTATCTTCTTATTTCTTCTTACCAAGATCTATATGATTCGTGTGAATCTTAT 637
Qy      187 TyrLeuCysArgAsnProPheAspThrPheIleSerSerTyrPheTyr---ThrAsnAsn 205
Db      638 TACATTTGACAGGAGCCCAAGATATCAATTTGTTCAATGTATCATTTATGACAGGAC 697
Qy      206 IleLysSerGluSerValSerProValLeuLeuAspGlnAlaPheAspLeuTyrCysArg 225
Db      698 TGCAAATTCCAAAATGCTCAACCGCTTCACTTATGAGAGGTTGGAATTTATGAA 757
Qy      226 GlyValIleGlyPheGlyProPheTyrGluHisMetLeuGlyTyrTyrArgGluSerLeu 245
Db      758 GGTGTAAGCCCGGTTTGGCGCTTATTTGGACCAATTTTGGGTTCTGGAAGCAAGTTTC 817
Qy      246 LysArgProGluLysValPhePheLeuArgTyrGluLysLeuLysAspAspIleGluThr 265
Db      818 GAACATCCGAGCAAGATTATGTTCTTGAAATATGAAAGAAATGAAGATATCTGTTTG 877

```



```
APPLICANT: JIN, HUA
APPLICANT: LAI, FANG MING
APPLICANT: LEROBERT, MARTIN
TITLE OF INVENTION: NUCLEAR FERTILITY RESTORER GENES AND METHODS OF USE IN
FILE REFERENCE: 16313-0210
CURRENT APPLICATION NUMBER: US/10/345,072
CURRENT FILING DATE: 2003-01-16
PRIORITY APPLICATION NUMBER: PCT/US02/22217
PRIORITY FILING DATE: 2002-07-12
PRIORITY APPLICATION NUMBER: 60/305,026
PRIORITY FILING DATE: 2001-07-12
PRIORITY APPLICATION NUMBER: 60/305,363
PRIORITY FILING DATE: 2001-07-13
PRIORITY APPLICATION NUMBER: 60/308,736
PRIORITY FILING DATE: 2001-07-30
NUMBER OF SEQ ID NOS: 179
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 87
LENGTH: 271990
TYPE: DNA
ORGANISM: Raphanus sativum
FEATURE:
NAME/KEY: modified base
LOCATION: (144241)..(144300)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-345-072-87

Alignment Scores:
Pred. No.: 1,136-67 Length: 271990
Score: 671.50 Matches: 139
Percent Similarity: 61.49% Conservative: 75
Best Local Similarity: 39.94% Mismatches: 122
Query Match: 35.38% Indels: 13
DB: Gaps: 5

US-10-019-931-3 (1-359) x US-10-345-072-87 (1-271990)

QY 15 SerpsermetCysGHSlySLeuGluLeuLysGluGlyThr----- 30
DB 244933 AACATATCTCTTTGTGAACGATTAACAGAACTCAAGAAACCAATGATCCG 244874
QY 31 ArgasrValProlYsAlaGluGluAspGluLysSerCysGluPheGluMetLeu 50
DB 244873 AAGGAGCTTCCTGGAACCTTGAAGACGACCAAGATTAAGTGAAGAACGAGATGTC 244814
QY 51 AspSerLeuProlYsGluArgGlyTrrPargThrArgTyrLeuPheGluGlyPhe 70
DB 244813 TCTTCGCTCTTCAACACAGATTAACAGGCGATTAAGATTTGAATATCAAGGATGT 244754
QY 71 TrpCysGlnAlaLysGluIleGlnAlaIleMetSerPheGlnLysPheGlnSerLeu 90
DB 244753 TGGATTATTATCAACACCTCTCAAGGTGTCTC-AATTTCCAGAGGCTTTCAACCGCA 244695
QY 91 GluAsnAspValIleLeuAlaThrIleProlYsSerGlyThrTrpLeuAlaIleu 110
DB 244694 GACCTAGTGAATCAATGCTTTCGATCCCAATCAAGGACCTACTGCTCAAGGCTCTC 244635
QY 111 ThrPheThrIleLeuAsnArgHisArgPheAspProValAlaSerSerThrAsnHisPro 130
DB 244634 ACACTCGCTCTGTTGAAGATCAAG-----AACCACTCTTTCGATCATCTCT 244587
QY 131 LeuPheThrSerAsnProHisAspLeuValProPhePheGluTyrLysLeuTyrAlaAsn 150
DB 244586 CTCTATATCAATATCCCATGAGCTTATACATCTTGGAGATCGATGTACCAAGCA 244527
QY 151 GlyAspValProAspLeuSerGlyLeu---AlaSerProGlnTrpPheAlaThrHisLeu 169
DB 244526 AGCTCAAGCTCTTAACCTAGGCAAGTTCACAGACTCCGAGGCTGTTCCTGCTACATG 244467
QY 170 ProPheGlySerLeuLysGluThrIleGluLysProGlyValLysValTyrLeuCys 189
DB 244466 CCACTGACACAGATCCAGAAAGCACTCAAGCACTCTCTTGCAGAGATTGTGAGGTGTGC 244407
```

```
QY 190 ArgAsnProPheAspThrPheIleSerSerTrpHisTyrThrAsnAlaIleYsSerGlu 209
DB 244406 AGGAACGTGAAGGACAGTTCATCTCTGTTGTTTACGCTGTCTATATATAATC 244347
QY 210 SerValSerProValLeuLeuAspGlnAlaPheAspLeuTyrCysArgGlyValIleGly 229
DB 244346 GAACCAACCAAGAGATTCCTGAGTCAATTTTAACAGATTTCTGCAGTAACCAACTAT 244287
QY 230 PheGlyProPheTrpGluHisMetLeuGlyTyrTrpArgGluSerLeuLysArgProGlu 249
DB 244286 TTTCGACCTTTTGGGATCATCTCTGATTAAGTTCGAGAGAACCTTGGAAACCAAG 244227
QY 250 LysValPhePheLeuArgTyrGluAspLeuLysAspAlaIleGluThrAsnLeuLysArg 269
DB 244226 CATCTCTTTCATGAGATTAAGAGATTAAGAAACCAAGCTCTGTGATCAATCAAGAA 244167
QY 270 LeuAlaThrPheLeuGluLeuProPheThrGluGluGluArgLysGlyValValLys 289
DB 244166 CTTCGGACCTTCTGGGATGTCCTTTTACTAAGCAAGAAAGATAGTGATCTGTGAC 244107
QY 290 AlaIleAlaGluLeuCysSerPheGluAsnLeuLysLysGluValAsnLysSerAsn 309
DB 244106 GGGATCTTGACCTCTGCTCTGCTGAATCTGACGAGTTGGAGCTTAACMAAACGGG 244047
QY 310 LysSerIleLysAsnPheGluAsnArgPheLeuPheArgLysGlyValLysSerAspTrp 329
DB 244046 --ACAATTAACCAATGAGGACCAACAGTTTTCCTGTAAGAGAGAGTGGTACTCG 243990
QY 330 ValAsnTyrLeuSerProSerGlnValGlu-ArgLeuSerAlaLeuValAspLeuLys 349
DB 243989 AAAAATATATCTTACG---TCTGAATGAGAAAGATGACATGATCAATCAAGAAACT 243933
QY 349 uGlyGlySerGlyLeuThrPhe 356
DB 243932 TCAAGGTTCTGTTGAGTTT 243911

RESULT 12
US-10-767-795-1750
; Sequence 1750, Application US/10767795
; Publication No. US20040181830A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongmei
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53534)B
; CURRENT APPLICATION NUMBER: US/10/767,795
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 1750
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Goseypium hirsutum
; OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C40_2
US-10-767-795-1750

Alignment Scores:
Pred. No.: 8,016-71 Length: 1251
Score: 666.50 Matches: 137
Percent Similarity: 58.19% Conservative: 69
Best Local Similarity: 38.70% Mismatches: 120
Query Match: 35.12% Indels: 28
DB: Gaps: 8

US-10-019-931-3 (1-359) x US-10-767-795-1750 (1-1251)

QY 16 PheSerMetCys---His-LysLeuGluLeuLeuLysGluGlyLysThrArgAspValPr 34
DB 40 TATAGTATTGCACTCATTAAGCTGTCTCATATGAAATCCACTTGAATTCATCCATGTTGA 99
```

34 olysalagiuglueapgluglyleusercyseryleuphenglumetleuaspserleupr 54  
100 GCAACATACGAGATGAG---TTTCAGAAATCTTCAAGAGAGATTTCTACTCTCC 156  
54 olysalagiuglytrp---ArgThrArgTyrLeuTyrleuphenglumetleuprCysgl 73  
157 TAAAGGGAAGAGCTGGGCTTCCGATCATCTATATCAAGATTTTGGTTGCTC 216  
73 nalaleuagluleglnalalemetserPheglumetleuphenglumetleuaspserleu 93  
217 CTCACCTCTTCCACAGAGATCTCTTGTAGCTCAACAGATTTGAGGCTCAACCCACTGA 276  
93 PValValleuAlaThrIleProlyserGlyThrThrTyrleuValaleuThrPheTh 113  
277 TATCATACTTTCTAGTGGCCCTAGAACGGCAGACCTGGTTAAATCTCTACTTCCG 336  
113 rtleuendnhrghlsarPheAspProValAlaSerSerThrAsnHisProleuPheTh 133  
337 TACTATTACAGAACTTCATACGAT-----GATTCACACCCCTTGTCTTTT 384  
133 rserAsnProHisAspLeuValProPhePhegluTyrlyseuTyrAlaAsnGlyAspVa 153  
385 CAGATGCCCTCATGATGTTGCTTTCATGAGCTT----- 421  
153 lProAspLeuSerGlyLeuAlaSerProArg-----ThrPheAlaThrHi 168  
422 ---GATCATGCTCATTTTCCGCTATTCGACATCTGGAAATTCCTTTTGGCCACTCA 477  
168 leuProPheglYserleuLysegluThrIleGlylybProGlyVallybValIlyle 188  
478 TGCTCTTATCTTCTTCTTACCTCAATATGATGATTCGTGGTGTGTAATTTATTAAT 537  
188 uCyArAsnProPheAspThrPheIleSerSerThrPheIlyr---ThrAsnAlly 207  
538 TTGCGAGGAGCCCAAGATACATTTGTTCAATGATATCAATCTGATACAGGTATGCCAA 597  
207 sSerGluSerValSerProValLeuLeuAspGlnAlaPheAspLeuTyrCyArGlyVa 227  
598 ATCCCAAAACACTCAACCATTCATGATGAGACGTTGCGATTTATTAAGAGGCGT 657  
227 lIleGlyPheglYProPheThrGlnHisMetleuIlyrTTPAArgGluSerleuLyAr 247  
658 AAGTTGGATGGGCTTATTTGGAGACATGTTGGGGTACTGGAAGCAAGCTTGAACA 717  
247 gProGlylybValPhePheLeuArgTyrGlyAspLeuLybAspAllyleuThrAsnle 267  
718 TCCAGACAAAGTTATGTTCTTGAATATGAGAACTGATGAAATACGTTTGTATCT 777  
267 uLyArGleuAlaThrPheleuGluLeuProPheThrGluGluGluArgLybGlyVa 287  
778 TAAAGAAATAGACAAATTTATGCTTCTTCGTCAGAGAAACAACAAGAGGGGT 837  
287 lVallybAlaIleAlaGluLeuCySerPheGluAsnleuLybLybLeuGluValAsnly 307  
838 GCCCTGAAAACATTTGACAGCTGTGCAATTTGAAGTTTAACTGCTTGAAGTAATA 897  
307 sSerAsnlySerIleLyAsn-----PheGluAsnArgPheLeuPheAr 322  
898 AACTGGGAACATCTCGACGGCAGAGAGATTTGGCGATGGAAATTAACATTTCTTCCG 957  
322 gLyAsGlyGluValSerAspTyrValAsnTyrleuSerProSerGlnValGlyArgLeuSe 342  
958 GAAGAGGAAGTTGAGACTGGAGAAATTAATTTGAACTGAATATGGCTGACGTTTGA 1017  
342 rAlaLeuValAspAspLybLeuGlylySerGlyLeuThr 355  
1018 CCACGAAACACTGCAAAAGTTTCGGCGTTCAAGTTTAAGT 1057

RESULT 13  
US-10-739-930-1663  
; Sequence 1663, Application US/10739930  
; Publication No. US20040216190A1  
; GENERAL INFORMATION:

APPLICANT: Kovalic, David K.  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
FILE REFERENCE: 38-21(5337)B  
CURRENT APPLICATION NUMBER: US/10/739, 930  
NUMBER OF SEQ ID NOS: 11088  
SEQ ID NO 1663  
LENGTH: 1358  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
US-10-739-930-1663  
US-10-739-930-1663

## Alignment Scores:

Pred. No.: 1.05e-70 Length: 1358  
Score: 666.00 Matches: 138  
Percent Similarity: 58.33% Conservative: 58  
Best Local Similarity: 41.07% Mismatches: 124  
Query Match: 35.09% Indels: 16  
DB: 8 Gaps: 8

US-10-019-931-3 (1-359) x US-10-739-930-1663 (1-1358)

34 ProlybAlagiuglueap---GluGly-----LeuserCysgluPheGln 47  
179 CCCAGCAAGAGAGCCAGATGCGGAGGAGGAGAAACGAACCTGAGCTCTTACAGATTTGCC 238  
48 GlumetleuAspSerleuProlygluArglyTyrPArgThrArgTyrleuTyrleuPhe 67  
239 AACTGTGCTCTCTCCCTCCAGCTCCAGATCCAAAGGCTTGTCCAAACAATTTACCGGCAC 298  
68 ---GlnGlyPheThrCyArgGlnAlaLysegluIleGlnAlaIleMetserPheGlnlybHis 86  
299 GACCAAGCTGACACTGACGCTGGTGGCATATGGGCGCCATATGTCGCGAGCGCTGC 358  
87 PheGlnSerleuGluAsnAspValValleuAlaThrIleProlyserGlyThrThrTyr 106  
359 TTCAAGCGCGCCCTCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 418  
107 LeuLybAlaLeuThrPheThrIleleuendnhrghlsarPheAspProVal-----Ala 124  
419 ATGAAGTCTCTCTCTGATGACGACCGTGCACCGG---AAGAGCGCTCCGCTGCGGCC 475  
125 SerSerThrAsnHisProleuPheThrSerAsnProHisAspLeuValProPheGlu 144  
476 GCCGCGACCAACATCTCTCAACAGCTGGGCGCCCGCCACGAATCATCAAGTTCTTGA 535  
145 TyrLybLeuTyrAlaAsnGlyAspValProAspLeuSerGlyLeuAlaSerProArgThr 164  
536 TACCAAGCTCTACAGCGCGGAGAGAGGCTCCGAGCTGACAGAGCTCCCGACCTTGGCTC 595  
165 PheAlaThrHisleuProPheGlySerleu---LybgluThrIleGluLybProGlyVal 183  
596 TTCCGAGACGACGCTCTTCTGTCGCTGCGAGCTCCGAGCTCCGAGCCAGCGGCGTGC 655  
184 LysValValTyrleuCyArgAsnProPheAspThrPheIleSerSerThrPheIlyrThr 203  
656 AAGATCGTCTACGCTGCGCGGAGCCCAAGACCACTCGTCTCTCTGAGACTTGCCTC 715  
204 AsnAsnle-----LysSerGluSerValSerProValLeuLeuAspGlnAlaPheAsp 221  
716 AAACAAATCTTCAGGGTCAAGAGATGGCTCGAGCTCTCTCGAGAGCGCGCGCGAG 775  
222 LeuTyrCyArgGlyValIleGlyPheGlyProPheThrGlnHisMetleuGlyTyrTyr 241  
776 CTGTCTCGCGCGCGCTGAGCCCGTTCCGGCGCTGACTGAGACCACTGCTGCGCTACCTGG 835  
242 ArgGluSerleuLybArgProGluLybValPhePheLeuArgTyrGlyAspLeuLybAsp 261  
836 CGCGCGCACCGCGCGAGACCCGAGCAAGGTGCTCTTCTTCAAGTACAGAGATGACAGAG 895

US-10-437-963-55524  
; Sequence 55524, Application US/10437963  
; Publication No. US20040123343A1

Alignment Scores:

US-10-019-931-3 (1-359) x US-10-437-963-55524 (1-1077)

[illegible]

RESULT 15  
US-10-425-115-167844  
; Sequence 167844, Application US/10425115  
; Publication No. US20040214272A1

```

: GENERAL INFORMATION:
: APPLICANT: La Rosa, Thomas J.
: APPLICANT: Kovalic, David K.
: APPLICANT: Zhou, Yihua
: APPLICANT: Cao, Yongwei
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21(5322)B
: CURRENT APPLICATION NUMBER: US/10/425,115
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ ID NOS: 369326
: SEQ ID NO 167844
: LENGTH: 1579
: TYPE: DNA
: ORGANISM: Zea mays
: FEATURE:
: OTHER INFORMATION: Clone ID: MRT4577 84655C.1

```

US-10-425-115-167844

## Alignment Scores:

Pred. No.:	1,91e-68	Length:	1579
Score:	648.50	Matches:	136
Percent Similarity:	55.84%	Conservative:	60
Best Local Similarity:	38.75%	Mismatches:	142
Query Match:	34.17%	Indels:	13
DB:	8	Gaps:	6

US-10-019-931-3 (1-359) x US-10-425-115-167844 (1-1579)

```
QY 14 ProSerPheSerMeCyHisIstLeuGluleuLeuLysGluGlyLysThrArg----- 31
DB 277 CCATGCTTCTCTGTAACCAAGCCATGGCGCAAGCTGCTCCGGCTCCAGCAAGATCGAC 336
QY 32 ---AspValProLysAlaGluGluGlyLeuSerCysGluPheGlnGluMetLeu 50
DB 337 ACCAAGCTCCACCTCAGCAGCAGCGCTCAACCCAGAGCTACGCGACCTGGTA 396
QY 51 AspSerLeuProLysGluArgGlyThrArgThrArgTyrlLeuPheGlnGlyPhe 70
DB 397 TCCACTCTCCCGCTAGAGCGGGTGGTATCG-----CTGTCGAGTACCAAGACTAC 450
QY 71 TrpCysGlnAlaLysGluLysIleGlnAlaIleMetSerPheGlnLysHisPheGlnSerLeu 90
DB 451 TGGCTGAACCCCGGAGAGCTCCAGAAACATCATCCCGTGAAGAGCTCACAAGCCTGCG 510
QY 91 GluAsnAspValValLeuAlaThrIleProLysSerGlyThrThrTrpLeuLysAlaLeu 110
DB 511 GCCGACGACATGCTCTCCGCCACTCACTGAATGCGGACACAGTGGCTGAAGGCGCTG 570
QY 111 ThrPheThrIleLeuAsnArgHisArgPheAspProValAlaSerSerThrAsnHisPro 130
DB 571 GCCTTGCCCATCAGACCCGCGCGCCACCAACAGCCTTCCGCGCGCCGACCAACCC 630
QY 131 LeuPheThrSerAsnProHisAspLeuValProPhePheGluTyrlLysLeuTyrlaAsn 150
DB 631 CTGCTCACCCTGTCACCCGAGAGAGCGCGCTGCGCACCTGAGGTGCACAGCGGGGCGAG 690
QY 151 GlyAspValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeuPro 170
DB 691 GGC---CTGGCGGACATCGAGAGGCTCGCTGCGGAGGCTGCTGCGGACGACCTGCC 747
QY 171 PheGlySerLeuLysGluThrIleGluLysProGlyValLysValValTyrlLeuCysArg 190
DB 748 CTCTCCCTGCTCCCGCGCGCGGCGCACCTCCGGCTGCGCGCTGCTGACTGTGCGG 807
QY 191 AsnProPheAspThrPheIleSerSerThrHisTyrlThrAsnAsnIleLysSerGluSer 210
DB 808 CAGCCCAAGGACGTGTGGTCTCGCTGCGCACTTGTGAAGGCGCATGCGCGG---GGC 864
QY 211 ValSerProValLeuLeuAspGlnAlaPheAspLeuTyrlCysArgGlyValIleGlyPhe 230
DB 865 AGGTCCCCCGTGAAGCTGAGCCGCGCTGCGCATGTTCTGCGAGGGGCTTCGCTTTC 924
QY 231 GlyProPheThrGluHisMetLeuGlyTyrlTrpArgGluSerLeuLysArgProGluLys 250
DB 925 GGGCCCGCTTGGAGGACACTTCACTCGAGTCTGGAAGAGAGCCTGCGAGGCTTGAACGG 984
QY 251 ValPhePheLeuArgTyrlGluAspLeuLysAspAspIleGluThrAsnLeuLysArgLeu 270
DB 985 GTCCTTTTCTCAGGATCAGAGGAGATGTGGCGGACCCGCTGCGTGCCTGAAGACGCTC 1044
QY 271 AlaThrPheLeuGluLeuProPheThrGluGluGluGluArgLysGlyValValLysAla 290
DB 1045 GCCGCTTCTTCCGCGCTCCGTTCAACGAGAGAGAGGCGGCTCCGAGAGAG 1104
QY 291 IleAlaGluLeuCysSerPheGluAsnLeuLysLysLeuGluValAsnLysSerAsnLys 310
DB 1105 ATTGTAGAGCTTGAGCTTCAAGATGTGAAGCGCTTGAAGACATCGAACCGA--- 1161
QY 311 SerIleLysAsnPheGluAsnArgPheLeu-----PheArgLysGlyGlu 325
```

```
DB 1162 GATCTTGATTCGGAGACACACGCGTGATCGAAATCTACCTTCTTCAGAGAGGCGAAG 1221
QY 326 ValSerAspTrpValAsnTyrlLeuSerProSerGlnValGluArgLeuSerAlaLeuVal 345
DB 1222 GTTGGGACTGGAGAACCACTTGACAAAGAGATGGCGCAAGAGTGAAGCGCGTTTTC 1281
QY 346 AspAspLysLeuGlyGlySerGlyLeuThrPhe 356
DB 1282 GAAGCAAGCTCCAGGATCTGTCTCGTGT 1314
```

Search completed: December 23, 2005, 22:02:17  
Job time : 959 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: December 23, 2005, 19:18:45 ; Search time 284 Seconds  
(without alignments)  
656.003 Million cell updates/sec

Title: US-10-019-931-3

Perfect score: 1898  
Sequence: 1 MATSSKSIIMAFPSMCH.....RLNALVDKKGSGLTRRLS 359

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4168288 seqs, 259477437 residues  
Total number of hits satisfying chosen parameters: 8336576

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xip  
-O=/cgn2\_1/USPRO.spool/p/US10019931/runat.16122005.165220.18532/app.query.fasta\_1.519  
-DB=published.Applications.NA.New -QFMT=fastap -SUFFIX=p2n.rnpn -MINMATCH=0.1  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10019931@CGN 1.183 @runat.16122005.165220.18532  
-NCPU=6 -ICPU=3 -NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100  
-LONGLOC -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPO=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications NA.New.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	274	14.4	1718	7	US-11-179-843-2
2	99	5.2	1027	6	US-10-750-185-42392
3	90	4.7	165857	7	US-11-121-086-34
4	87.5	4.6	2412	6	US-10-821-234-791
5	86.5	4.6	909	6	US-10-793-626-1143
6	86.5	4.6	3315	6	US-10-793-626-13820
7	86.5	4.6	3431	6	US-10-793-626-4435
8	86.5	4.6	3431	6	US-10-793-626-4435

9	85.5	4.5	1316	6	US-10-750-185-55045	Sequence 55045, A
10	85	4.5	171433	7	US-11-121-086-85	Sequence 85, Appl
11	84.5	4.5	67088	7	US-11-117-187-186	Sequence 186, Appl
12	84	4.4	420	6	US-10-821-234-822	Sequence 822, Appl
13	84	4.4	142303	7	US-11-121-086-41	Sequence 42, Appl
14	83.5	4.4	1281	6	US-10-750-185-31307	Sequence 31307, A
15	82.5	4.3	2112	7	US-11-094-586-1	Sequence 1, Appl1
16	82.5	4.3	3312	6	US-10-793-626-3465	Sequence 3465, Ap
17	82	4.3	2442	6	US-10-793-626-3843	Sequence 3843, Ap
18	82	4.3	3084	6	US-10-793-626-3105	Sequence 3105, Ap
19	81.5	4.3	3148	6	US-10-793-626-3105	Sequence 3105, Ap
20	81.5	4.3	2069	6	US-10-689-742-139	Sequence 139, Appl
21	81.5	4.3	156297	7	US-11-121-086-65	Sequence 65, Appl
22	81	4.3	7292	6	US-10-750-185-42616	Sequence 42616, A
23	81	4.3	1169	7	US-11-112-908-514	Sequence 514, Appl
24	80.5	4.2	1549	6	US-10-995-561-18	Sequence 18, Appl
25	80.5	4.2	1943	6	US-10-995-561-17	Sequence 17, Appl
26	80.5	4.2	109974	7	US-11-117-187-204	Sequence 204, Appl
27	80.5	4.2	161994	7	US-11-112-908-57	Sequence 57, Appl
28	80.5	4.2	16866	7	US-11-112-908-59	Sequence 59, Appl
29	80.5	4.2	170285	7	US-11-112-908-58	Sequence 58, Appl
30	80.5	4.2	2250	6	US-10-793-626-137	Sequence 137, Appl
31	80	4.2	2705	6	US-10-750-185-50240	Sequence 50240, A
32	80	4.2	2950	6	US-10-793-626-3349	Sequence 3349, Ap
33	80	4.2	4198	6	US-10-928-446A-1	Sequence 1, Appl1
34	80	4.2	108000	6	US-10-928-446A-181	Sequence 181, Appl
35	80	4.2	108000	6	US-10-928-446A-181	Sequence 181, Appl
36	80	4.2	108000	6	US-10-928-446A-183	Sequence 183, Appl
37	80	4.2	108000	6	US-10-928-446A-185	Sequence 185, Appl
38	80	4.2	108000	6	US-10-928-446A-187	Sequence 187, Appl
39	80	4.2	108000	6	US-10-928-446A-189	Sequence 189, Appl
40	80	4.2	108000	6	US-10-928-446A-191	Sequence 191, Appl
41	80	4.2	108000	6	US-10-928-446A-193	Sequence 193, Appl
42	80	4.2	108000	6	US-10-928-446A-195	Sequence 195, Appl
43	80	4.2	108000	6	US-10-928-446A-197	Sequence 197, Appl
44	80	4.2	108000	6	US-10-928-446A-199	Sequence 199, Appl
45	80	4.2	108000	6	US-10-928-446A-199	Sequence 199, Appl

ALIGNMENTS

RESULT 1  
US-11-179-843-2  
Sequence 2, Application US/11179843  
Publication No. US20050277153A1  
GENERAL INFORMATION:  
APPLICANT: Thomae, Bianca A.  
APPLICANT: Wieben, Eric D.  
APPLICANT: Weinshilboum, Richard M.  
TITLE OF INVENTION: Sulfoltransferase SULT2A1 Sequence  
FILE REFERENCE: 07039-276001  
CURRENT APPLICATION NUMBER: US/11/179, 843  
CURRENT FILING DATE: 2005-07-12  
PRIOR APPLICATION NUMBER: US/10/177, 570  
PRIOR FILING DATE: 2002-10-01  
PRIOR APPLICATION NUMBER: US 60/300, 165  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 1718  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-11-179-843-2  
Alignment Scores:  
Pred. No.: 1.12e-21  
Score: 274.00  
Percent Similarity: 46.24%  
Best Local Similarity: 27.24%  
Query Match: 14.44%  
DB: 7  
Length: 1718  
Matches: 76  
Conservative: 53  
Mismatch: 118  
Indels: 32  
Gaps: 9





TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES  
 FILE REFERENCE: 09138.6000-00000  
 CURRENT APPLICATION NUMBER: US/11/121.086  
 CURRENT FILING DATE: 2005-05-04  
 PRIOR APPLICATION NUMBER: 60/567,570  
 PRIOR FILING DATE: 2004-05-04  
 NUMBER OF SEQ ID NOS: 107  
 SOFTWARE: PatentIn version 3.3  
 SEQ ID NO 34  
 LENGTH: 165857  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-11-121-086-34

Alignment Scores:  
 Pred. No.: 9.17e+03 Length: 165857  
 Score: 90.00 Matches: 75  
 Percent Similarity: 30.82% Conservative: 27  
 Best Local Similarity: 22.66% Mismatches: 95  
 Query Match: 4.74% Indels: 134  
 DB: 7 Gaps: 18

US-10-019-931-3 (1-359) x US-11-121-086-34 (1-165857)

Db 26 LysGluGlyLysThrArgAspValProLysAlaGluGlyLeuSerCysGlu 45  
 8825 AAGCAGGGTGTGTCATTTTCTACCAATTCACAGCACTCAGCAATGGCTTTCAA 8884  
 46 PheGlnGluMetLeuAspSerLeuProLysGluArgGlyTrpArgTyrLeuTyr 65  
 8885 TTTCAG-----TATATTGACCTTCCTCACAATTAGGGTGAAG----- 8923  
 66 LeuPheGlnGlyPheTrpCysGlnAlaLysGluLeuAlaIleMetSerPheGlnLys 85  
 8924 ---TACACTACATTTTGGTTTAACAACAACAAAGCAATCCCTCCACCCCAAAA 8980  
 86 HisPheGlnSerLeuGluLysAspValValLeuAlaThrIlePro-----LysSerGly 103  
 8981 TAAATTT-----TATTTAGGAATTAATCTCTCTTAATAATCTGGC 9019  
 104 ThrThrTrp----- 106  
 9020 ATGTTTGGAGGCAAGAAAGTATGATTAAGAAACACACAAGCAACAGAAACT 9079  
 107 -----LeuLysAlaLeuThrPheThrIleLeuAsnAlaGlyHisArgPhe----- 120  
 9080 TTCATGTAGCCAAAGTTCTCACACACCACATCTCTGAATGCACTTATATCTCAGGG 9139  
 121 -----AspProValAlaSerSer-----ThrAsnHisProLeu 131  
 9140 AGACAAAGTAAGGTAATCTATATCAAGGCTATGAGGAGCTCCAAAAAATCATAGTA 9199  
 132 PheThrSerAsnProHis-----AspLeuValProPhePheGlu 144  
 9200 CAAGTTCAAAACAATTAACAGGTTTGAAGGCTTTGACAACGTTCCACCTTCACTCA 9259  
 145 TyrLysLeuTyrAlaAsnGlyAspValPro----- 154  
 9260 TGGAGACTTGTAGCTTAACACAATGTCCCATATATCTGCTGANTGTTTCCAGCACTAA 9319  
 155 -----AspLeuSerGlyLeuAlaSerProArgThrPhe 165  
 9320 AACCTGGAACAATGTTAAAAAGAAAAAATTAAGTGGGGTGGG-----ATC 9367  
 166 AlaThrHisLeuProPheGlySerLeuLysGluThr----- 177  
 9368 TTAGAGTAGATCCATTCAGAGGATGACACAGAACTTAAGTTAAATAAGAGACTCA 9427  
 178 ---IleGlyLysPheGlyValLysValValLysLeuCysValArgAsnProPheAsp----- 194  
 9428 AAATCCAGAGCCCTTAGGGTTCCAGCATGACAAATCCAAAG---CTTTTTCATTTCTCG 9484  
 195 -----ThrPheIleSerSer---TrpHisTyrThrAsnAlaIleLysSerGluSer 210

Db 9485 TTGCTCCACATTCATCTGCTTGG----- 9514  
 211 ValSerProValLeuLeuAspGlnAlaPheAspLeuTyrCysArgGlyValIleGlyPhe 230  
 9515 ---CTTCAAGAAACATGACATCTACTTACAACTTACAGTGGGCTCAGCATTACCATAA 9571  
 231 GlyProPheTrp-----GluHisMetLeuGlyTyrTrp---ArgLysSerLeu--- 245  
 9572 GATCATTTGACTCAATTTTAAAGATTTTGGGGGGGGTGGAGAGGAATTTCTTTTGG 9631  
 245 ----- 245  
 9632 AGCAGTCTAATAATATAAGATGATTTGTACTGTTTCTTGTGACACCTT 9691  
 246 -----LysArgProGluLysValPhePhe 253  
 Db 9692 CCAAGAAACAAAAAGAAAGAAATCTTTT 9724

RESULT 4  
 US-11-121-086-83/C  
 ; Sequence 83, Application US/11/121086  
 ; Publication No. US20050266459A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: POULSEN, TIM S.  
 ; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES  
 ; FILE REFERENCE: 09138.6000-00000  
 ; CURRENT APPLICATION NUMBER: US/11/121.086  
 ; CURRENT FILING DATE: 2005-05-04  
 ; PRIOR APPLICATION NUMBER: 60/567,570  
 ; PRIOR FILING DATE: 2004-05-04  
 ; NUMBER OF SEQ ID NOS: 107  
 ; SOFTWARE: PatentIn version 3.3  
 ; SEQ ID NO 83  
 ; LENGTH: 187745  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-11-121-086-83

Alignment Scores:  
 Pred. No.: 1.11e+04 Length: 187745  
 Score: 90.00 Matches: 63  
 Percent Similarity: 34.45% Conservative: 40  
 Best Local Similarity: 21.07% Mismatches: 113  
 Query Match: 4.74% Indels: 83  
 DB: 7 Gaps: 11

US-10-019-931-3 (1-359) x US-11-121-086-83 (1-187745)

Qy 64 LeuTyrLeuPheGlnGlyPheTrpCysGlnAlaLysGluLeuAlaIleMetSerPhe 83  
 123825 CTCTTCCTTCAGGGGAGAAAGCCCAAGCAGGC----- 123790  
 84 GlnLysHisPheGlnSerLeuGluLysAspValValLeuAlaThrIleProLysSerGly 103  
 123789 -----CTTCTCAACCTTTTACACCCACTGG 123763  
 104 ThrThr-----TrpLeuLysAlaLeuThrPhe----- 112  
 123762 GTCAACAGCATCTTCTCTGCTGAGGGGCTTGGGGAAGCCAAACCAACCCCTGCCCC 123703  
 113 -----ThrIleLeuAsnArgHisArgPheAspProValAlaSerSerThrAsnHisPro 130  
 123702 TTCCCAACCTTGGCAGGACACAGATCACTGATGCTGCTCATCTGGCAATTCCC 123643  
 131 LeuPheThrSerAsnProHisAspLeuValProPhePheGluTyrLys-LeuTyrAlaAs 150  
 123642 AAGCAGCAGGAGATCTACGAGCAGACAGTGTCTTGGGCAAGCCCTTGAAGCCAC 123583  
 150 nGlyAspValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeuP 170  
 123582 TGCTCC-----CCAGGGCTCTGAAAGCAAGAAATGAAGTGGTGAAGCTGAG 123535

QY	170	oPhelgYserLeuLygluThrlleGluySvProglYVallyValValYrlleuCySAr	190
Db	123534	CAACGGAGAAAGTGAAGGGCTCCCTGGAGAAAGCTGGTGGAAAG-----CACAAAATGCCG	123491
QY	190	GAAnPrOPheArThrPheIlseSerTrpHisTyTrtHAsnAllelySerGluse	210
Db	123480	CTTCTGTGCCAAAGTATTGGCAGT-----	123456
QY	210	rValSerProValLeuLeuAerGlInAlpheArleTyTrCyArGlyValIlleGlyPh	230
Db	123455	-----GACAGTCCCTGGCAGATCCACCTTCGTTCACACGGGTGA	123415
QY	230	eGlYProPheTrpGlInHleMerleuGlYTrTrArGluSerleuLyArProglY	250
Db	123414	GAGGCCATTAAAGTCAATGTCTGTGAAC--CGTTTACACCCCGTGGCAACTCA	123358
QY	250	eValPhePheLeuArGlyTrGluArleuLySArAerPlleGluThrAsn-----	266
Db	123357	AAATGCAATTTCCACCGGCACTCGTAGAAAGTAAACCAATGTGCAATGAACCAACCCAGT	123298
QY	267	-----leuLySArGleuAlArThrPheleuGluleProPhe-----	278
Db	123297	ACCAGAGCACTAGACTATGTCTATTCCAGAGTGGCTTGCCTTAATGTATGTCCGAGCC	123238
QY	279	-----ThrlgluGluy-----GluArGlyGlyA	287
Db	123237	ACCAGAAAGGCCGAGAGAGAGGCAACCCACTCAGGTGAGGGGTTGAGCGCAAGCCTCT	123178
QY	287	lVallyAlaIleAlaGluLeuCySerPheGlAenLeuLyAlaGluAlaAnly	307
Db	123177	GGTGGCTTCACAACAGCACTAGTGCACAGAGAGCTGACTGTCTCCACCAAGTGC	123118
QY	307	sSerAnlySerIleLyAsnPheLu-----AsnArGPhleuPheArGlyS	323
Db	123117	AGGCACAGCCAGGCTCCAGACTCCCTGTTCAATAGTTGTGCTCATGA	123063
RESULT 5			
US-10-821-234-791			
Sequence 791, Application US/10821234			
Publication No. US20050255114A1			
GENERAL INFORMATION:			
APPLICANT: Labat, Ivan			
APPLICANT: Stache-Grain, Birgit			
APPLICANT: Andarmani, Susan			
APPLICANT: Tang, Y. Tom			
TITLE OF INVENTION: Method for Diagnosis and Treatment of Precipalsia			
FILE REFERENCE: 821A			
CURRENT APPLICATION NUMBER: US/10/821,234			
CURRENT FILING DATE: 2004-04-07			
PRIOR APPLICATION NUMBER: US 60/462,047			
PRIOR FILING DATE: 2003-04-07			
NUMBER OF SEQ ID NOS: 1704			
SOFTWARE: pc_seq_genes version 1.0			
SEQ ID NO 791			
LENGTH: 2412			
TYPE: DNA			
ORGANISM: Homo sapiens			
US-10-821-234-791			
Alignment Scores:			
Pred. No.: 25.6 Length: 2412			
Score: 87.50 Matches: 87			
Percent Similarity: 34.70% Conservative: 57			
Best Local Similarity: 20.96% Mismatches: 138			
Query Match: 4.61% Indels: 133			
DB: 6 Gaps: 24			
US-10-019-931-3 (1-359) x US-10-821-234-791 (1-2412)			
QY	23	GIuLeuLeuLygluGluySerTrArGArPValProLyAlaGluGluAerGluyLeu	42
Db	871	GAGCCCATAGAGAAAGAAAGCAAGCCAAAGAAAGAAAGAAAGAAATCTGATGTAGTGAAGCT	930

OY		43	SerCygluPheGlnImeLeuAspSerLeuProLygGluAArgGly-----	58
Dd		931	GAGTAAAG---GAAGAAGAAAGAAAGAAACCAAGCTAAAAAGTTGAAAACCT	987
OY		59	---TrpAlaThrArgTyrLeuTyrLeuPheGlnIyPheTrpCysGln--AlaLeuGlu	76
Dd		988	GTCGTGGAGCTGGAAACTATGATATCATCAACCATTATGGCAGAACCCATAAGAA	1047
OY		77	IleGln-----AlaIleMetSerPheGlnLysHisPheGlnInsrLeuGluAsnAspVal	94
Dd		1048	GTAAGAAGAGATGAATACAAGCTTTCTTACAAATCAATTT---TGAAAGAAAGTAGAAC	1104
OY		95	ValIeuAlaThrIle-----ProLyserGlyThrTrkTrpleuLysAlaIleuThrpHe	112
Dd		1105	CCCATGGCTTATATCTACTTACGTGGAGGGGAAGTTACCTTCAATATCAATTTATTT	1164
OY		113	----ThrIleLeuAsnArgHisArgPheAspProValAlaSerSer-----	126
Dd		1165	GTACCCACATGTGCTCCACGCGTGTGTGTGCAGAAATGATGATCTAAAGAGCGAATAC	1224
OY		127	-----ThrAsnHisProLeuPheThrsSerAsnProHisAspLeuValProPhe	142
Dd		1225	ATTAAAGCTATATGCGCGCGGTATTCATCATCACAGACACTTCCATGATATGATGCT--	1281
OY		143	PheGlnTyrLysLeuTyrAlaAsnGly-----AspValProAspLeuSerGlyLeuAla	160
Dd		1282	---AAATACCTCAATTTTGTGCAGAGGtGTGGTGAGTCACAGATGATCTCCCCTGAATGTT	1338
OY		161	SerProArgTrpThrPheAlaThrHisLeuProPheGlySerLeuLysGluThrIleGluLys	180
Dd		1339	TCCCGCAGACTCTTCCAGCAACAT-----AAAGCTGTTAAAGTCATTAAGAG	1386
OY		181	ProGlyValLys-----ValValTyrLeuCysArgAsnProPhe--Asp	194
Dd		1387	AAGCTGTGTCTTAAGCGCTGAGATGATCATGAAGATGCTGATGATTAATACAAATGAT	1446
OY		195	ThrPhe-----	196
Dd		1447	ACTTTTGTGAAGAAATTGGTACCAACATCAAGCTGTGTGATTTGAAGACCACTCGAAT	1506
OY		197	-----LseSerTrpHisTyrThrAsnAsnLys	207
Dd		1507	CGAACAGCTCTGCTAAACTCTTAAGTCCAGTCTTCTCATCTCATCCAACATGACATTACT	1566
OY		208	SerGluSerValSerProValLeuLeuAspGlnAlaPheAspLeuTyrCysArgGlyVal	227
Dd		1567	AGC-----CTAGACCAG-----	1578
OY		228	IleGlyPheGlyProPheTrpGluHisMetLeuGlyTyrTrpArgGluSerLeuLysArg	247
Dd		1579	-----TATGTGGAAGATGACAGAAAA	1602
OY		248	ProGluLysValPhePheLeuArgTyrGluAspLeuLysAspAspIleGluThrAsnLeu	267
Dd		1603	CAGACAAAAATCTACTTATGCGCGGTCCACAGCAAGAAA---GAGGCTGAATCTTC--	1656
OY		268	LysArgLeuAlaThrPheLeuGluLeuProPheThrIugluGluGluArgLysGly---	286
Dd		1657	-----CCATTTGTTAGCGACTTCTGAAAAAGGGCTAT	1689
OY		287	-----ValValLysAlaIleAlaGluLeuCys-----	295
Dd		1690	GAACTTATTAACCTACACAGAACTGTGGATGAATACTGTATTACAGGCCCTTCCGAATTT	1749
OY		296	-----SerPheGluAsnLeuLysLysLeuGluValAsnLysSerAsnLysSerIle	312
Dd		1750	GATGGAAAGAGGTTCCAGAAATGTTGGCCAG---GAGAGAGTGAAGTTCCAGTGAAGTAGAG	1806
OY		313	LysAsnPheGluAsnArgPheLeuPheArgLysGlyLysGlyValasrAspTrpValAsnTyr	332
Dd		1807	AAAACCTAAGAGAGATCGTGAAGACAGTTGAGAA---GAAATTGAGCCTCTGTGAATATGG	1863

Qy 333 LeuserProSerGlnVal-----GluArgLeuSer 342  
Db 1864 ATGAAAGATAAAGCCCTTAAGACAGATTGAAAAGCGTGTGCTCTCAGCGCTGACA 1923  
Qy 343 -----AlaLeuValAspAspLysLeuGlyGlySerGly 353  
Db 1924 GAATCTCCGTGTGCTTGTGTGGCCAGCAGCATGAGATGCTGTGC 1968  
RESULT 6  
US-10-793-626-1143  
; Sequence 1143, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P03480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1143  
; LENGTH: 909  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
US-10-793-626-1143  
Alignment Scores:  
Pred. No.: 7.39 Length: 909  
Score: 86.50 Matches: 67  
Percent Similarity: 37.79% Conservative: 46  
Best Local Similarity: 22.41% Mismatches: 87  
Query Match: 4.56% Indels: 99  
DB: Gaps: 18  
US-10-019-931-3 (1-359) x US-10-793-626-1143 (1-909)  
Qy 10 PrometAlaIleProSerPheSerMetCysHisIleLysLeuGluLeuGlyLys 29  
Db 130 CCTGTACTATTACTTAATTT-----GTAGAAATTGATGGCTCAAGAAAT 174  
Qy 30 ThrArgAspValProLysAlaGluGluAspGlu---GlyLeuSerCysGluPheGlnGlu 48  
Db 175 -----CCTTCATTGCCAAAAGTACCAATTGCAATAGATTTAAGAGTAAAGAA 222  
Qy 49 MetLeuAspSerLeuProLysGluArgGlyTTPArgThrArgTyrLeuTyrLeuPheGln 68  
Db 223 TTATTTAAATCTCTA----- 237  
Qy 69 GlyPheTyrCysGlnAlaLysGluIleGlnAlaIleMetSerPheGlnLysPheGln 88  
Db 238 -----GTTAAAGAAATTGATATTATTAAGT----- 264  
Qy 89 SerLeuGluAsnAspValValLeuAlaThrIleProLysSerGlyThrThrTyrLeuLys 108  
Db 265 -----CGTATGACGTGATTTTAGATGTTATG-----TTTTTAA 300  
Qy 109 AlaLeuThrPheThrIleLeuAsnArgHisArgPheAspProValAlaSerSerThrAsn 128  
Db 301 GCTTAAACTGAAAAATTATTTACGTATTA-----GAATCAAGAGAGCG 348  
Qy 129 HisProLeuPheThrSerAsnProHisAspLeuValProPhePheGluTyrLysLeuTyr 148  
Db 349 CACCCACTAAATGAACAAGACAGATCATTAATAGTGCATTA----- 393  
Qy 149 AlaAsnGlyAspValProAspLeuSerGlyLeuAlaSerProArgThrPhe----- 165  
Db 394 ---AATGAGAAAGCGAAGCATCTATCAGAAATCCGAGTATCGTAATATAGCATGAT 450

Qy 166 AlaThrHisLeuProPheGlySerLeuLysGluThrIleGluLysProGlyValLysVal 185  
Db 451 ACAACAAATTTAAACCTTAAGATTAAACCAACGCTTTCAAG----- 495  
Qy 186 ValTyrLeuCysArgAsnProPheAspThrPhe-----IleSerSerTyrPheTyr 202  
Db 496 TTTATTTATGATGAAGAAC---TTTGAACAATTATCAATCAACGAGCAAGATTTGGCTTTC 552  
Qy 203 ThrAsnAsnIleLysSerGluSerValSerProValLeuLeuAspGlnAlaPheAspLeu 222  
Db 553 AAGCATGTATTAACAATGATGCT-----GATTTAGTTTGAAGTTC 594  
Qy 223 -----TyrCysArgGlyValIleGlyPheGlyProPheTyrGlu 235  
Db 595 AGATTTTACTTAATCCCTACTATGTAGAGAAATGGCTCATTTACTGTTAGATGAG 654  
Qy 236 HisMetLeuGlyTyr-----TTPArgGluSerLeuLysArgProGluLysValPhe 252  
Db 655 CCAGTGTACATTAACGTTATGAAGTGAAGAAAGAAC-----CAAAATATT 699  
Qy 253 PheLeuArgTyrGluAspLeuLysAspAspIleGluThrAsnLeuLysArgLeuAlaThr 272  
Db 700 TTTCATTAATTAACAAGATTTA-----TTAAATATTATG----- 732  
Qy 273 PheLeuGluLeuProPheThrGluGluGluArgLysGlyValLysAlaIle 291  
Db 733 -----ATTCTGTGCTACAAAAAGAAAGGTAATTCGCAATTGTTATGCTATA 780  
RESULT 7  
US-10-793-626-3820  
; Sequence 3820, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P03480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3820  
; LENGTH: 3315  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
US-10-793-626-3820  
Alignment Scores:  
Pred. No.: 55 Length: 3315  
Score: 86.50 Matches: 67  
Percent Similarity: 37.79% Conservative: 46  
Best Local Similarity: 22.41% Mismatches: 87  
Query Match: 4.56% Indels: 99  
DB: Gaps: 18  
US-10-019-931-3 (1-359) x US-10-793-626-3820 (1-3315)  
Qy 10 PrometAlaIleProSerPheSerMetCysHisIleLysLeuGluLeuGlyLys 29  
Db 540 CCTGTACTATTACTTAATTT-----GTAGAAATTGATGGCTCAAGAAAT 584  
Qy 30 ThrArgAspValProLysAlaGluGluAspGlu---GlyLeuSerCysGluPheGlnGlu 48  
Db 585 -----CCTTCATTGCCAAAAGTACCAATTGCAATAGATTTAAGAGTAAAGAA 632  
Qy 49 MetLeuAspSerLeuProLysGluArgGlyTTPArgThrArgTyrLeuTyrLeuPheGln 68  
Db 633 TTATTTAAATCTCTA----- 647

```
QY 69 GlyPheTrpCySGlnAlaIySGluIleGlnAlaIleMetSerPheGlnIlyHisPheGln 88
DB 648 -----GTTAAAGAAATTTGATTATTAAAGT----- 674
QY 89 SerLeuGluAenAspValValLeuAlaThrIleProIlySerGlyThrThrTrpLeuIys 108
DB 675 -----CGTAAAGACGTGATTTTAAGATGTTATG----- 710
QY 109 AlaLeuThrPheThrIleLeuAenArgHisArgPheAspProValAlaSerSerThrAsn 128
DB 711 GTTAAAACTGAAAAATTTTTCACGTTATTA-----GAATCAAGAAAGAGG 758
QY 129 HisProLeuPheThrSerAsnProHisAspLeuValProPhePheGluIlyTrpIlyLeuTr 148
DB 759 CACCCACTTAATGAAACAAGACAAAGATCATTAATATGATGCAATA----- 803
QY 149 AlaAenGlyAspValProAspLeuSerGlyLeuAlaSerProArgThrPhe----- 165
DB 804 ---AATGAGGAACGTGAACATCTATCAGAAATCCGAAAGATGCGTAAATTACGTGATTGAT 860
QY 166 AlaThrHisIleuProPheGlySerLeuIySGluThrIleGluIyProGlyValIlyVal 185
DB 861 ACAACAAATTTAAACCTTAAGATTAAAGCAACGATTTCAAAG----- 905
QY 186 ValIlyLeuCyArgAsnProPheAspThrPhe-----IleSerSerTrpHisTr 202
DB 906 TTTTATTATAGATAAAC---TTTGAACATTATTACAATCAACGTGACAAAGTTTCGGTTTC 962
QY 203 ThrAsnAenIleIySerGluSerValSerProValLeuLeuAspGlnAlaPheAspLeu 222
DB 963 AAGCATGTGATCAAAATGATGCT-----GATTAGTTTGTATGATGC 1004
QY 223 -----TyrCyArgGlyValIleGlyPheGlyValIleGlyPheGlyProPheTrpGlu 235
DB 1005 AGATTTCACCTTAATCCCTACTATGTAGAGAAATTCGCTCATTTACTGTTAATGATGAG 1064
QY 236 HisMetLeuGlyTrp-----TrpArgIleuSerLeuIyAspArgProGluIyValPhe 252
DB 1065 CCAAGTCAATTAATGATTATGAAAGTGAAGAAAGAAC-----CAATATATT 1109
QY 253 PheLeuArgTrpGluAspLeuIyAspAspIleGluThrAsnLeuIyAspGluAlaThr 272
DB 1110 TTTGATTAATTAACAGATTTA-----TTAAATTTTATG----- 1142
QY 273 PheLeuGluLeuProPheThrGluGluGluIyAspGlyValIyValIyAlaIle 291
DB 1143 -----ATTCTGGCTACAAAAAGAGTAATGCCAATTGGTTATTGCTATA 1190

RESULT 8
US-10-793-626-4435/c
; Sequence 4435, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4435
; LENGTH: 3431
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4435

Alignment Scores: 58 Length: 3431
Pred. No.:
```

```
Score: 86.50 Matches: 67
Percent Similarity: 37.79% Conservative: 46
Best Local Similarity: 22.41% Mismatches: 87
Query Match: 4.56% Indels: 99
DB: 6 Gaps: 18

US-10-019-931-3 (1-359) x US-10-793-626-4435 (1-3431)
QY 10 PrometAlaIleProSerPheSerMetCyHisIlyLeuGluLeuIySGluGlyIys 29
DB 1598 CCGTACACTTAATCACTTAATTT-----GTAGAAATGATGCGTCAAGAAAT 1554
QY 30 ThrArgAspValProIyAlaGluGluAspGlu---GlyLeuSerCySGluPheGlnIly 48
DB 1553 -----CCTTCATTGCAAAAGTACCAATTCGAATGATTAATTAAGGTAAAGGA 1506
QY 49 MetLeuAspSerLeuProIyGluArgGlyTrpArgThrArgTrpIlyLeuPheGln 68
DB 1505 TTATTTAATCTCTA----- 1491
QY 69 GlyPheTrpCySGlnAlaIySGluIleGlnAlaIleMetSerPheGlnIlyHisPheGln 88
DB 1490 -----GTAAAGAAATTCATATTAATAAGT----- 1464
QY 89 SerLeuGluAenAspValValLeuAlaThrIleProIlySerGlyThrThrTrpLeuIys 108
DB 1463 -----CGTAAATGACGTGATTTTATGATGTATG-----TTTTTAGAA 1428
QY 109 AlaLeuThrPheThrIleLeuAenArgHisArgPheAspProValAlaSerSerThrAsn 128
DB 1427 GCTAAAACTGAAAAATTAATTTTACGTTATTAATA-----GAATCAAGAAAGACG 1380
QY 129 HisProLeuPheThrSerAsnProHisAspLeuValProPhePheGluTrpIlyLeuTr 148
DB 1379 CACCCACTTAATGAAACAAGCAACAAAGATCATTAATATGATGCAATA----- 1335
QY 149 AlaAenGlyAspValProAspLeuSerGlyLeuAlaSerProArgThrPhe----- 165
DB 1334 ---AATGAGGAACGTGAACATCTATCAGAAATCCGAAATGCGTAAATTACGTGATTGAT 1278
QY 166 AlaThrHisIleuProPheGlySerLeuIySGluThrIleGluIyProGlyValIlyVal 185
DB 1277 ACAACAAATTTAAACCTTAAGAAATTTAAAGCAACGATTTCAAAG----- 1233
QY 186 ValIlyLeuCyArgAsnProPheAspThrPhe-----IleSerSerTrpHisTr 202
DB 1232 TTTTATTATAGATAAAC---TTTGAACATTATTACAATCAACGTGACAAAGTTTCGGTTTC 1176
QY 203 ThrAsnAenIleIySerGluSerValSerProValLeuLeuAspGlnAlaPheAspLeu 222
DB 1175 AAGCATGTGATCAAAATGATGCT-----GATTAGTTTGTATGATGC 1134
QY 223 -----TyrCyArgGlyValIleGlyPheGlyProPheTrpGlu 235
DB 1133 AGATTTCACCTTAATCCCTACTATGTAGAGAAATTCGCTCATTTACTGTTAATGATGAG 1074
QY 236 HisMetLeuGlyTrp-----TrpArgIleuSerLeuIyAspArgProGluIyValPhe 252
DB 1073 CCAAGTCAATTAATGATTATGAAAGTGAAGAAAGAAC-----CAATATATT 1029
QY 253 PheLeuArgTrpGluAspLeuIyAspAspIleGluThrAsnLeuIyAspGluAlaThr 272
DB 1028 TTTGATTAATTAACAGATTTA-----TTAAATTTATG----- 996
QY 273 PheLeuGluLeuProPheThrGluGluGluIyAspGlyValIyValIyAlaIle 291
DB 995 -----ATTCTGGCTACAAAAAGAGTAATGCCAATTGGTTATTGCTATA 948

RESULT 9
US-10-750-185-55045
; Sequence 55045, Application US/10750185
; Publication No. US2005026063A1
; GENERAL INFORMATION:
```

```

; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: PARTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55045
; LENGTH: 1316
; TYPE: DNA
; ORGANISM: Bovine 1986880686798
US-10-750-185-55045

Alignment Scores:
Pred. No.: 17.2 Length: 1316
Score: 85.50 Matches: 40
Percent Similarity: 36.65% Conservative: 19
Best Local Similarity: 24.84% Mismatches: 50
Query Match: 4.50% Indels: 52
Gaps: 6

US-10-019-931-3 (1-359) x US-10-750-185-55045 (1-1316)
QY 116 AaahrgHsaArgpheaSPProValAlaSerThrAenHsIProLeuPheThSerAen 135
Db 846 AATTTCATTTGTTGACCCGCTCCACCTGTCACAGCTCACTGCTCACCCAGTAAC 905
QY 136 ProHsaPleuValProPhePheGluTyrIleuValAlaAenGlyAspValProAsp 155
Db 906 CCTGAGGGCTCTGTTCACTTTTC-----CTTCCC----- 935
QY 156 LeuSerGlyLeuAlaSerProArgThrPheAlaThrHsIleuProPheGlySerIleuLys 175
Db 936 -----ACACAGCATACACTGACAGCTGGGCTTTCCAAA 971
QY 176 GluThrIleGlyLysProGlyValLysValValTyrLeuCyArgAsnProPheAspThr 195
Db 972 CCACCACTCGAACACTTCAACGAAATGATCATATAT----- 1007
QY 196 PheIleSerSerTPrHisTyrThrAenIleLysSerGluSerValSer-ProValIle 215
Db 1008 -----CAAGCCATCTT 1019
QY 215 uLeuAspGlnAlaPheAspLeuTyr---CyArgGlyValIleGlyPheGlyProPheTr 234
Db 1020 GATGCCAGAAACCTTGACATTTGCTCCGCGACGGAATGACAACTTCATCTGCCAAGCTG 1079
QY 234 PGIuHisMetLeuGlyTyrTPrArgGluSerLeuLysArgProGlyLysValPhePheLe 254
Db 1080 GACTCATTTGCGCGGC-----AAAGTGAAGGACAAATTTTCACTTC 1121
QY 254 uArgTyr-----GluAspLeuLysAspAspIleGluThrAenLeuLysArgLeuAlaTh 272
Db 1122 ACTCATCTCGAGCAACAATTCAAGACACATTAATAACCTTAAGCGTTAAGCATTC 1181
QY 272 x 272
Db 1182 T 1182

RESULT 10
US-11-121-086-85/c
; Sequence 85, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.

```

```

; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 85
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-85

Alignment Scores:
Pred. No.: 3,78e+04 Length: 171423
Score: 85.00 Matches: 78
Percent Similarity: 34.69% Conservative: 58
Best Local Similarity: 19.90% Mismatch: 13
Query Match: 4.48% Indels: 114
DB: 7 Gaps: 16

US-10-019-931-3 (1-359) x US-11-121-086-85 (1-171423)
QY 20 HislyleuGluleuLeuLygluLy-----LysThrArgAspValProLys 35
      |||||
      :::::
Db 21819 TACAAACTGGAAACCTAGAAAGATGATGATGATTCCTGACACATCTCTGTCACAA 21760
QY 36 AlagluGluAspGluLy-----LeuSerCysGluPheGluGlu 48
      |||||
      :::::
Db 21759 CTGAACCAAGAAAAAATCTGATTTCTGAAACAGACCAATATAGCTGTAATTAATCA 21700
QY 49 MetLeuAspSerLeuProLysGluArg-----GlyTrpArgThrArgTyr 63
      |||||
      :::::
Db 21699 GTAGTAAATAGCTTACCAACCAAAAAATGTCAGTACCTGATGATTAACACTAAATTC 21640
QY 64 LeuTyrLeuPheGluGlyPheTyrCysGlnAlaLysGluIleGlnAlaIleuSerPhe 83
      |||||
      :::::
Db 21639 TACCAAGATGACAAA-----AGAGAGCTACTACATTCCTTA----- 21604
QY 84 GlnLysHisPheGlnSerLeuGluAsn-----AspVal 94
      |||||
      :::::
Db 21603 CAGAAACTATTTCCAAAAATGAAAGAAAGGAGCTCCCACTCATCTGTGATGGCC 21544
QY 95 ValLeuAlaThrIleProLysSerGly-----ThrThrIleuLysAlaLeuThrPhe 112
      |||||
      :::::
Db 21543 ATCATCTTCTGTGATACCAAACTGCGACAGACACACAAAAAGAAAACATTATAGTC 21484
QY 113 ThrIleLeuAsnArgHisArgPheAspProValAlaSerSerThrAsnHisPro----- 130
      |||||
      :::::
Db 21483 AATATCTTATATGACATCATGATGGAAAAACCTCAATATAAATATTGCMAAACCAACA 21424
QY 131 -----LeuPheThrSerAsnProHisAspLeuValPro----- 141
      |||||
      :::::
Db 21423 TCAAAATTGACATCAAAAAGCTAAATTCCACCAATCAAGAGAGCTCATCCCTGGGATC 21364
QY 142 -----PhePheGluTyrLysLeuTyrAlaAsn-----Gly 151
      |||||
      :::::
Db 21363 CAAAGTTGGTTCAACATACACAAATATATATAGTCGATTAATCACATTAACGAAGCTAA 21304
QY 152 AspValProAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPhe 171
      |||||
      :::::
Db 21303 GACAAAACCAACATGATTAATCTCAATATGACGAAAAGAGCTTCAATATAAATTCAACATC 21244
QY 172 GlySerLeuLysGluThrIleGluLysProGlyValLysValValTyrLeu----- 188
      |||||
      :::::
Db 21243 AATTCATTGAAAA--ACTGCATATAAATAGGATTAATCGAGTATATCTTCAAAATATA 21186
QY 189 -----CysArgAsnProPheAspThrPheIleSerSerTyrHisIleTyrThrAsn 205
      |||||
      :::::
Db 21185 AGAGCCATCTGTGACAAACCCGACCAACTTATATCTGAATGGGCAAAATCTGAGAGCA 21126

```

QY 206 IleySerGIuser-----ValSerProValIleuLeuAspGln 218  
Db 21125 TTCCCTTGAATCTGGCAGCATGCAAGAGGCCCTCTCTCACTCTCTATTTCAAAATA 21066  
QY 219 AAlaPheAspLeuTyrcysArgGlyVal----- 227  
Db 21065 GTATTGAAAGTCCAGCCAGCAGCTCAAGTAAGAAATAAAGGCATCCAAATA 21006  
QY 228 -----IleGlyPheGlyProPheTrpGluHisMetLeuGlyTyrTrpArgGlu 243  
Db 21005 GGAAGAGAGAAAGTCAAGCTATCTCTGTTTGGACATCAATATCTATATCTAAGAAAC 20946  
QY 244 SerLeuLysArgProGluLysValPhePheLeuArgTyrGluAspLeuLysAspIle 263  
Db 20945 CCCATAGTCTCAGCTCAAAAGCTCTCATCTGATTAATCACTCAGCAAAAGTTTGGGATAC 20886  
QY 264 GluThrAsnLeuLysArgLeuAlaThrPheLeuGlu----- 275  
Db 20885 AAAATCAATGTACAAATAACACTTAACGTTCTATACACCAACAGCCAGCTGAGAGC 20826  
QY 276 -----LeuProPheThrGluGluGluArg----- 284  
Db 20825 CAAATCAGAAATTAAATCCCATTCACAAATGGCCACAAAGAAATAAATACCTAGAAATAC 20766  
QY 285 -----LysGlyValValLysAlaIleAlaGluLeuCysSerPheGlu 298  
Db 20765 AGCTAATCAGAAAGTGAAAGTTTCTCATCTGAGATCAACAACACTGCTCA----- 20712  
QY 299 AsnLeuLysLysLeuGluValAsnLysSerAsnLys 310  
Db 20711 -----AGAAATCAGAGATTAACACAAACAAATAGA 20682  
RESULT 11  
US-11-117-187-186  
; Sequence 186, Application US/11117187  
; Publication No. US2005026560A1  
; GENERAL INFORMATION:  
; APPLICANT: PREUSS, DAPHNE  
; APPLICANT: COPENHAVER, GREGORY  
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS  
; FILE REFERENCE: ARCD:309US  
; CURRENT APPLICATION NUMBER: US/11/117,187  
; CURRENT FILING DATE: 2005-04-28  
; PRIOR APPLICATION NUMBER: US/09/531,120  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/125,219  
; PRIOR FILING DATE: 1999-03-16  
; NUMBER OF SEQ ID NOS: 212  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 186  
; LENGTH: 67088  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (27227)..(64728)  
; OTHER INFORMATION: N = A, C G, or T/U  
US-11-117-187-186  
Alignment Scores:  
Pred. No.: 1,01e+04 Length: 67088  
Score: 84.50 Matches: 94  
Percent Similarity: 31.00% Conservative: 52  
Best Local Similarity: 19.96% Mismatches: 159  
Query Match: 4.45% Indels: 167  
DB: 7 Gaps: 19  
US-10-019-931-3 (1-359) x US-11-117-187-186 (1-67088)  
QY 3 ThisSerMetLysSerIleProMetAlaIleProSerPheSerMetCysHisLysLeu 22  
Db 18187 ACATTCTCGTACGCTCTCTCACTGACCAACATCATCACTGACATCTCTCTCAACGCGCTC 18246

QY 23 GluLeuLeuLysGluGlyLysThrArgAspValProLysAlaGluGluAspGlyLeu 42  
Db 18247 TCTTTCACTGAGAAACGCTCTCTCTCTATCTCTCACTCACTCACTCACTCACTCACT 18306  
QY 43 SerCysGluPheGluGluMetLeuAspSerLeuProLysGluArgGlyTyrArgThrArg 62  
Db 18307 TCACGCTCAAGCGCTCTCTCTCACTGACCTCGCTCCCACTCAACCATTCCTCCT 18366  
QY 63 TyrLeuTyrLeuPheGluGlyPheTrpCysGluAlaLysGluIleGluAlaIleMetSer 82  
Db 18367 -----CACCACTTCGTAACTTCTCTCACTGACCAAAA-----TTCAAGC 18405  
QY 83 PheGluLysHis-----PheGluSerLeuGluAsnAspVal--- 94  
Db 18406 TTTCACTGCTCAGCGCACTGCTCTCTCTCTCTCTCACTCAAGACGAGACCGGTTTCA 18465  
QY 95 -----ValLeuAlaThrIleProLysSerGlyThrThrTyrLeuLys 108  
Db 18466 TCATCTCAACCGCTCTGCTCTCTCTCACTCAAGCGCAAGACCACTCACTCACT 18525  
QY 109 AlaLeuThrPhe-----ThrIleLeuAsnArgHisArgPheAsp----- 121  
Db 18526 CCGATTCTTTTCTCACTCTCTCACTCACTCACTCACTCACTCACTCACTCACTCACT 18585  
QY 122 -----ProValAlaSerSerThrAsnHisProLeuPheThrSer 134  
Db 18586 CTCTTTTACTGACCGCGGACCGGCTCTCACTCACTCACTCACTCACTCACTCACTCA 18645  
QY 135 AsnProHisAspLeuValProPhePheGluTyrLysLeuTyrAlaAsnGlyAsp-ValPr 154  
Db 18646 CCTGCGCACTCATTTG-----CGCCTCGTTTATCTCTCACTGACCGCTCTCAAAAC 18659  
QY 154 OAspLeuSerGlyLeuAlaSerProArgThrPheAlaThrHisLeuProPheGlySerIle 174  
Db 18700 GCCACTGCTCTCTCTCACTCTCGCGTTTCACTCG-ACACGCGATTACCGTCTCACTTCG 18758  
QY 174 uLysGluThrIleGlyLysProGlyValLysValValTyrLeuCysArgAsnProPheAs 194  
Db 18759 TCTTCACTCACTCGCTAA-----CTCGCA 18782  
QY 194 PThrPheIleSerSerTrpHisTyrThrAsnAsnIleLysSerGluSerValSerProVa 214  
Db 18783 AACCATTAATTCACGTAC-----TCGACCGT 18809  
QY 214 IleuLeuAspGlnAlaPheAsp-----LeuTyrCys-ArgGlyValIleG 229  
Db 18810 AATACTGACCGGTGACTGACCGGTTAGTTTGCATTATTTGACTTAACATATTGA 18869  
QY 229 LysPheGly-----ProPhe-----TrpGlu- 235  
Db 18870 CGTTTGCTTTGAGTTACATTTCTTTTCAGGAAATCAATATGATGATGATGATGATG 18929  
QY 236 -----HisMetLeuGly-Tyr-----Trp----- 241  
Db 18930 TCTTCACTGATGCGGATTAACAATGATGATGATGATGATGATGATGATGATGATG 18989  
QY 241 ----- 241  
Db 18990 AGAGAGCGACAAGCTTATGAGAGCTTCAGAGCGGAGACCAACGCTCACTGAAGC 19049  
QY 242 -----ArgGluSerLeuLysArgProGluLysValPhePheLeuArgTyrGluAsp 258  
Db 19050 AATGAGAGAGAGCTGAGATTGCTAGAGAAAGAGACAAATGACGACGAGATGAGTTG 19109  
QY 259 LeuLysAspAspIle----- 263  
Db 19110 ATCGACGATGATTAATGACGTGAGTATGAGCTGATGATGATGATGATGATGATG 19169  
QY 264 -----GluThrAsnLeuLysArgLeuAlaThrPheLeuGluLeu----- 276  
Db 19170 TTGAACAGCGCTAATGAAGTCCAGTGAAGAGATCAATCAACTTTTGAAGCTGAAGAC 19229  
QY 277 -----ProPheThrGluGluGluArgGlyVal----- 287

[illegible]

```

Db      226 GAGAAATGCTACTATTATTTACCTTTGAAGATGGCAA----- 261
Oy      269 ArgIeuAlaIthrPheIeuGluIeuProPheIthrGluIeuGluValArgLysGlyValVal 288
Db      262 -----CATTT----- 267
Oy      289 LysAlaIleAlaGluIeuCysSerPheGluAlaLeuLysLysLysGluValAlaAsnLysSer 308
Db      268 -----GAGCTGTGCATCTATATGTGCGCTCAACAGGAATACAGTAATGTGTAAT 315
Oy      309 AsnLysSerIleLysAsnPheGluAlaAsnArgPheLeuPheArgLysGlyGluValSerAsp 328
Db      316 GGGCAAGCGCATTTACACATTTGCCCATCATTC----- 348
Oy      329 TrpValAsnIleLysSerProSerGluAlaGluArgLysSerAlaLeuValAsnAspLys 348
Db      349 -----CCGCCAGATCTGTGAAGATGCTGCAGATCTTCAGAGATATCTCC 393
Oy      349 Leu 349
Db      394 CTG 396

RESULT 13
US-11-121-086-42/c
; Sequence 42, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 42
; LENGTH: 142303
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-42

Alignment Scores:
Pred. No.: 3,73e+04 Length: 142303
Score: 84.00 Matches: 84
Percent Similarity: 33.41% Conservative: 53
Best Local Similarity: 20.49% Mismatches: 114
Query Match: 4.43% Indels: 159
DB: 7 Gaps: 22

US-10-019-931-3 (1-359) x US-11-121-086-42 (1-142303)
Oy      4 SerSerMetLysSerIleProMetAlaIleProSerPheSerMetCysHisLysLeuGlu 23
Db      95742 AGTTCACGCCACGACAAAGTTAAATGTTCTTCCACAGAACAGTCCTGCTCAGAAAATGGA 95683033
Oy      24 LeuLeuLysGluGlySerThrArgAspValProLysAla-----GluGluAsp 39
Db      95682 TGAACGACGAGGACGCGTGTGAGCCGCGCTGTAATCCACGAATTTTGGAGGCGCGAG 95623333
Oy      40 GluGlyLysSerCysGluPheGlnIleMetLeuAspSerLeuProLysGluArgLysTyr 59
Db      95622 ATGGGAGAGATCACTTAGAGTCAGAGATTTGAGACCAAGCCCTGGCAACA-----TGG 95572727
Oy      60 Arg-----ThrArgIleLeuTyrLysPheGlnGlyPheTyrCys----- 72
Db      95571 CAAAACTGTCTCTCTACTTAAAAATACAAAATTAAGCTAGACATGTGTGATGCTCTGA 95512
Oy      73 -----GlnAlaLysGlu-----Ile 77
Db      95511 ATTCACAGCTACTAGAGGCTGTGAGAGGAGCATTCCTTGAGCCCAAGAGGACGAGGCTTG 95452

```

Oy	78	GLNALIlelmetSerPheGlnlyuVhIshPheGlnSerleu---	GLuShnSerValValleu	96
Db	95451	CAGTGACCCGAGATTTGATCCAAATGCACATCCAGCGGGGGTAAAGAGGAGAACCTTATCTCA		95392
Oy	97	AlaThrIleProlyuSerGly---ThrThrTyrLeuValAlaLeuThrPheThrIleLeu	115	
Db	95391	AAAGAAATTAAGAAATGGAATGGAAGAACCCCTGGCTGCAGTACTGCACC-----CTA	95341	
Oy	116	AsnArgHisArgPheAspProVal-AlaSerSerThrAsnHisProLeuPheThSerAs	135	
Db	95340	CGGCGCTTTTATATGATGATGATGATGATCTTATCTCTAAGGACACATTTATTACTAGAAC	95281	
Oy	135	nProhIAspLeuValIProPhePheGlnlyuTyrlyuLeuTyrAlaAsnGlyAspValProAs	155	
Db	95380	AAAGTATGAATTAAGAACCATTTA-----AAATCTTCTGTATTTGGAGTCTGAGGTG	95230	
Oy	155	pluSerGlyLeuAlaSerProArgThrPheAlaThrHis-	168	
Db	95329	GGTGGGGGAGATTGCTTGAGCCCGAGAGTATGAGAACAGCTGGGCAACATATGGGACCC	95170	
Oy	168	-----	168	
Db	95169	CATCTTATTAATAAAAAATCTTCTGGAATTAAGCAATGAGAGCATTAATAGAAAAGT	95110	
Oy	169	-LeuProPheGlySerleuylsGlnThrIleGlyuSProGlyValIlyValValTyrIle	188	
Db	95109	GTTCCTCCCTAAGACACTCTGACTGGCTGGGTGAAGATGAGGGGTAGAAGGTGCAGAGAT	95050	
Oy	188	uCyArgAsnProPhe-----AspThrPheIleSerSerTyrHis-	201	
Db	95049	C---AGACTACACTCTCTCGGAAACTTCCCTGCTCTCACACACACTGCTTTTCTGAC	94939	
Oy	201	-----	201	
Db	94992	TGACTTGTGTGACAGGTGGTATGTTACAGCATTAATAGTTCTAGCTCAGTTAACTACATCTC	94933	
Oy	202	-----TyrThr-----	203	
Db	94932	GGGCTCAGTAACTACACACTCAGGGTGGTCTTATTTCTGCGGGGGCGGGAGTGAAC	94873	
Oy	204	---AsnAsnIleuysSerGlnSerValSerProValIleuLeuAsp-----G1	218	
Db	94872	AGAGATATGCTTAAGATTTTCCAGGTTCTCCCTATATTTGTCTGATGCCAGAAATTCGAC	94813	
Oy	218	nAlaPheAspLeuTyrCys-----ArgGlyVal-----I1	228	
Db	94812	AGCATTTACACTCTTGTGTCACATTGAAGTAAACATTCTCCCAAGAGGATTATGTTGAG	94753	
Oy	228	eGlyPheGlyProPhe-----TyrGlnHisIleLeuGlyTyrTrrPar	242	
Db	94752	CGGCTTTGGAGTTTCTTACAGGGAGTGAAGGTGGGAACAT-----	94710	
Oy	242	gGluSerLeuylsArgProGlnlyuSValPhePheLeuArgTyrGlu-----	257	
Db	94709	---AAGAGCCTGGGCAAAATGTGAGATTAAGGAAAGAAACGACCGAAGAGG	94666	
Oy	258	---AspLeuylsAspAspIleGlyThrAsnLeuylsArgLeuAlaTh	272	
Db	94659	GCTTTCTTTACTGTGAATATGTTGTGATGACATTAAGAC-----AC	94618	
Oy	272	rPheLeuGlnLeuProPheThrGlnGlu	281	
Db	94617	ATTCTTAGAC-----TGACACGAGCG	94596	

```

1  APPLICANT: ROSENFIELD, David
2  APPLICANT: HOLM, Tom
3  APPLICANT: BATES, Stephen
4  APPLICANT: FANTIN, Dennis
5  TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
6  FILE REFERENCE: MW11100-2
7  CURRENT APPLICATION NUMBER: US/10/750,185
8  CURRENT FILING DATE: 2003-12-31
9  PRIOR APPLICATION NUMBER: US 60/437,482
10  PRIOR FILING DATE: 2002-12-31
11  NUMBER OF SEQ ID NOS: 64922
12  SOFTWARE: PatentIn version 3.1
13  SEQ ID NO 31307
14  LENGTH: 1261
15  TYPE: DNA
16  ORGANISM: Bovine
17  US-10-750-185-31307
18  198668891270493

```

Alignment Scores:	
Pred. No.:	28.6
Score:	83.50
Percent Similarity:	40.24
Best Local Similarity:	26.04
Query Match:	4.40%
DB:	6
Length:	128
Matches:	44
Conservative:	24
Mismatches:	58
Indels:	13
Gaps:	10

US-10-019-931-3 (1-359) x US-10-750-185-31307 (1-1281)

OY		15 SerPheSerMetCysHisAlaIleuInduIleuLys-----	26
Dd		969 AGCAGCAGCCCTTCGCATCTCTCAACTACTACGTGGCTCAAGTATTAAAGAATCTGCC	910
OY		27 -----GluGIlyLeuThrArgAspValProIlyValagiIuIuAerGIuGIly-----	41
Dd		909 TGCAGATGCAGAGACTCACGTTTCGATTCCTGGCTGGGGAAGATCCCTCGAGAAAGGAAAT	850
OY		42 -----LeuSerCysgluIuPhegiIuIuIeIuAerSerIuProIySGIuArgGIly	58
Dd		849 GGCACACTACTCCAGTGITCTTGCTGTGAAGATCTC-----CTGGAACAGAGAGCTGG	796
OY		59 Trp-----ArgThrArgTyIleuTytyIleuPhegiIuGIly	69
Dd		795 TGCGGCTAACAGTCCGCGGTGTCAAAAGATGGAGTACGATGAGCTGTTCCTTCAATAG	726
OY		70 PheTrpCysglnAlaIylsgluIleuIla-----IleMetSerPhegiIuIyHisIaPhe	87
Dd		735 GTCTGGCCAAAGCATGAGATGTGCTGCTCTTTTCCCTCCATTTTAAGTAGCACTTTC	676
OY		88 ---GlnSerIeuGlubnaArValValIleuAlaThrIleProIySerGIlyThrThrTrp	106
Dd		675 TGACTGCTTCCGGCTGTGATGTCTTTTA-----ATTCCAGGCCACGCCGGGCACTCT	622
OY		107 LeuIlyValIleuThrPheThrIleIeuIuAnArgHilaVapHeaerProvalAlaIaSerSer	126
Dd		621 CTGAACCATCTGACTCTGTACATGTGCTCCCTTGACAT-----CCAGTTCTTCCAGA	571
OY		127 Thr-----AsnHisProIeuPheThrSerAsn-----ProHiaAsp	138
Dd		570 TC CATCTGCAGACATCTCGTTCGTTCCGTGATTCCTGGTATTCTGTATGATGACCTGTGCC	511
OY		139 LeuValProPhePhegiIuTytyIyIsIeu	147
Dd		510 TCTGTCCCTTTCTTGATGCACAAAAGTT	484

RESULT 14  
US-10-750-185-31307/c  
; Sequence 31307, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard

RESULT 15  
US-11-094-586-1  
Sequence 1, Application US/11094586  
Publication No. US20050273886A1  
GENERAL INFORMATION:  
APPLICANT: Allen, Stephen M.  
APPLICANT: Butler, Karlene H.  
APPLICANT: Carlson, Thomas J.  
APPLICANT: Hiltz, William D.



; APPLICANT: Stoop, Johan M.  
 ; TITLE OF INVENTION: Placidic Phosphoglucosyltransferase Genes  
 ; FILE REFERENCE: B01451 USCIP  
 ; CURRENT APPLICATION NUMBER: US/11/094,586  
 ; CURRENT FILING DATE: 2005-03-30  
 ; PRIOR APPLICATION NUMBER: US 09/906,209  
 ; PRIOR FILING DATE: 2001-07-16  
 ; PRIOR APPLICATION NUMBER: US 60/218,712  
 ; PRIOR FILING DATE: 2000-07-17  
 ; NUMBER OF SEQ ID NOS: 32  
 ; SOFTWARE: Microsoft Office 97 & PatentIn Version 3.3  
 ; SEQ ID NO: 1  
 ; LENGTH: 2112  
 ; TYPE: DNA  
 ; ORGANISM: Typha latifolia  
 ; US-11-094-586-1

Alignment Scores:  
 Pred. No.: 81.5 Length: 2112  
 Score: 82.50 Matches: 55  
 Percent Similarity: 30.98% Conservative: 33  
 Best Local Similarity: 19.37% Mismatches: 87  
 Query Match: 4.35% Indels: 109  
 DB: 7 Gaps: 12

US-10-019-931-3 (1-359) x US-11-094-586-1 (1-2112)

Qy 114 ILEuAnAIGHsARPhesProValAlaserSerThraenHieProleuPheThr 133  
 Db 1108 ATCTTGAGAACGTTCTTT-----GTACACCATCAGATTCTGTGCAATGATGCC 1161  
 Qy 134 SerAenProHIsAerleuValProhePhegluTyryleuTyAlaenGlyAspVal 153  
 Db 1162 GCTAATGACAGCGCGCTATTCTTATTCCAA----- 1194  
 Qy 154 ProAerleuSerGlyleuAlaserProArghrPhealathrHieleuPro---PheGly 172  
 Db 1195 -----GCTGGTCCCAAGGACTGTGAGTGTATGCCAACAAGCGGT 1236  
 Qy 173 SerleuYsgluThrIleGluYerProGlyValYsValTyryleuCyarGAsnPro 192  
 Db 1237 GCTCTTGATCGTGTAGCCGAAATTTGAACCTT-----CCA 1272  
 Qy 193 PheAerPthrPheHieserSerTrpHisrTyrrhrAenHieYserGluSerValSer 212  
 Db 1273 TTTCTTGAGTTCCAACTGTTGAAAGTTTGTGAAATCTGATGATGCTGGAAAGTTG 1332  
 Qy 213 ProValleuAerPglAlaPheAerleuTyryCyarGlyValIleGlyPheGlyPro 232  
 Db 1333 TCCATCTGTGGGAGGAAGTTT-----GCCACAGTTCT 1368  
 Qy 233 -----PheTrpGluHieMetleuGly 239  
 Db 1369 GATGACATCCCGGAGAGAGATGACATCTGCGCTGTTTGGCTTGGCTTCCATAATGCG 1428  
 Qy 240 TyrrhrPargGluSerleuYsArgProGluYsVal----- 251  
 Db 1429 TACAGAAACAAAGACAAAGATTGGAGAAATTAGTCTCTGTTGAAGATATTGCTAAG 1488  
 Qy 252 -----PhePheleuArGlyrGluAerleuYsAerPsr 262  
 Db 1489 GAGCACTGGGCAAAATATGCAAGAACTTCTTTCTCGATATGAT----- 1533  
 Qy 263 IleGluThrAenleuYsArgleuAlathrPheleuGluLeuProPheThrGluGluGlu 282  
 Db 1534 -----TACGAAGAATGC 1545  
 Qy 283 GluArgYsglyValValIleAlaGluLeuCyserPheGluAenleuYsArg 302  
 Db 1546 GAATCGAAGAGCAAAATTAATGATGACGACCTTAGGACTTT----- 1590  
 Qy 303 LeuGluValAenleuYsAerAsnleuSerIleYsAenPheGluAenArgPheleuPheArg 322

Db 1591 ATCTGACAAAGCAAGCCTGGAGAACAAATATGAAATTATACTTCAATT----- 1641  
 Qy 323 LysGlyGluValSerAerTrpValAsnTyryleuSerProSerGluValGluArgleuSer 342  
 Db 1642 -----TCAGATGACTTTTCTTACACTGACCT-----GTAGAC 1674  
 Qy 343 AlaLeuValAerAerPlyrleuGly-----GlySerGlyleuThr 355  
 Db 1675 GGCAGTGTAGCATCCAAAGCAAGGCTACGATTGTTTTCACAGATGATCAAGGCTTATC 1734  
 Qy 356 PheArgleuSer 359  
 Db 1735 TATCCTCTCTCG 1746

Search completed: December 23, 2005, 22:08:33  
 Job time : 489 secs

**THIS PAGE BLANK (USPTO)**